TRANSACTIONS

OF THE

AMERICAN INSTITUTE

OF

ELECTRICAL ENGINEERS

EASTMAN KODAK CO.
RESEARCH LIBRARY
KODAK PARK

VOLUME 78
1959

PART II. APPLICATIONS AND INDUSTRY

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1960 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1959 (volume 78) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. (Part III is divided into two parts: III–A, consisting of pages 1–976; III–B, consisting of pages 977–1870. The Index for both parts appears at the end of III–B.) The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics	Part II. Applications and Industry	Part III. Power Apparatus and Systems
Aircraft and Flight Test Instrumentation	Aero-Space Transportation	Insulated Conductors
Basic Sciences	Chemical Industry	Power Generation
Broadcasting	Domestic and Commercial Applications	Power System Communications
Communication Switching Systems	Electric Heating	Protective Devices
Communication Theory	Electric Welding	Relays
Computing Devices	Feedback Control Systems	Rotating Machinery
Data Communication	General Industry Applications	Substations
Electrical Insulation	Industrial and Commercial Power Systems	Switchgear
Electrical Techniques in Medicine and Biology	Industrial Control	System Engineering
Electronic and High Frequency Instruments	Industrial Power Rectifiers	Transformers
Electronics	Land Transportation	Transmission and Distribution
Indicating and Integrating Instruments	Marine Transportation	
Magnetic Amplifiers	Metal Industry	
Nucleonic and Radiation Instrumentation	Mining Industry	Research
Nucleonics	Petroleum Industry	Safety
Radio Communication Systems	Production and Application of Light	Standards
Recording and Controlling Instrumentation		
Semiconductor Metallic Rectifiers		
Solid State Devices		
Special Instruments and Auxiliary Apparatus		
Telegraph Systems		
Telemetering		
Wire Communication Systems		

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 78 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., February 1–6, 1959.
2. Southeast and South Central District Meeting, Atlanta, Ga., April 8–10, 1959.
3. East Central District Meeting and Rubber and Plastics Conference, Akron, Ohio, April 22–24, 1959.
4. Empire District Meeting, Syracuse, N. Y., April 29–May 1, 1959.
5. Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959.
6. Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959.
7. AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959.
8. Fall General Meeting, Chicago, Ill., October 11–16, 1959.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Applications and Industry Index for 1959

1. Technical Subject Index

Absorber Geometry for a Solar Generator, Optimum Reflector-. Stineman.....332–7
A-C Electrical System Designs, Computer Evaluation of High-Temperature Aircraft. Sollecito, Swann.....434–44; disc. 444
A-C Generator for High-Temperature Electrical Systems, A Solid Rotor. Bateman..400–03; disc. 403
A-C Power Using Variable Speed Generation, Constant Frequency. Jessee, Spaven....411–16; disc. 417
A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances, Analysis of. Lind, Schmitz.....476–9; disc. 479
Adaptive Control Systems, Use of a Mathematical Error Criterion in the Design of. Merriam...506–12
Adaptive Systems, Executive-Controlled. Staffin.....523–30
Adjustment, Control by Stochastic. Bertram...485–91
Air Filters, Electrified Fibrous. Thomas, Woodfin.....276–8
(Air Transportation) A Solid Rotor A-C Generator for High-Temperature Electrical Systems. Bateman.....400–03; disc. 403
(Air Transportation) An Elementary Design Discussion of Thermoelectric Generation. Bollmeier.....445–50
(Air Transportation) Constant-Frequency Variable-Speed Frequency-Make-up Generators. Hoard.....297–301; disc. 301
(Air Transportation) Development of Fuses and Terminals for High-Temperature Applications. Bonwitt, Buttner.....267–72
(Air Transportation) Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems. Rias.....25–30
(Air Transportation) Nuclear Radiation and Electronic Instruments. Crittenden.....423–6
(Air Transportation) Optimum Reflector-Absorber Geometry for a Solar Generator. Stineman..332–7
(Air Transportation) Practical Considerations of an Ion Propulsion System. Lennert....326–31; disc. 331
(Air Transportation) Solar-Powered Thermoelectric Generator Design Considerations. Schuh, Tallent.....345–52
(Air Transportation) Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components. Sisco, Fiskin.....352–6
(Air Transportation) Temperature Prediction in Thermal-Lag Equipment. Richards.....462–5
(Air Transportation) Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications. Duncan, Rectanus....16–19
(Air Transportation) The Effect of Variable High-Altitude Humidity on the Wear of Nondusting Brushes. Moberly, Johnson.....263–7
(Air Transportation) The Electrical Analog for Determining Temperature Distribution Electrical Components. Eckert, Hartnett, Irvine, Birkebak..5–10
(Air Transportation) Variable-Speed Constant-Frequency Devices: A Survey of the Methods in Use and Proposed. Owen.....321–5; disc. 325
Aircraft A-C Electrical System Designs, Computer Evaluation of High-Temperature. Sollecito, Swann.....434–44; disc. 444
Aircraft Control Equipment, Functional Cycling to Assure Reliability of. Hulsey, Kessler.....356–60
Aircraft Electric Power Systems, Characteristics and Measurement of Ripple in. Markowitz.....418–23; disc. 423
Aircraft, Generator Insulation Systems Development for Hypersonic. Penn, Balke, Precopio.....255–9
Aircraft Generators, Reliability Analysis for. Duane, Yeager.....426–34
Aircraft, New Materials for Transformer Rectifier Unit in High-Temperature. Hoop, McIlvaine..450–6
Aircraft, Variable-Speed Constant-Frequency Generator System for. Chirgwin, Stratton.....304–09; disc. 309
Aluminum Bus for Shipboard Application. Thompson, Behr.....239–47; disc. 247
Aluminum-Copper Joints, The Effect of Elevated Temperature on Flash-Welded. Dixon, Nelson.....491–5; disc. 495
Amplifiers, Transient Response and the Stabilization of Feedback. Mulligan.....495–503
Amplistat-Regulated Sectional Drive at Southland Paper Mills, Transistor. Vickery, Shaad.....369–74; disc. 374
Analog Computer, Predicting the Performance of a Wind Tunnel Regulating System Using an. Herbst, Nichols, Keay, Miller.....248–52
Analog for Determining Temperature Distribution in Electrical Components, The Electrical. Eckert, Hartnett, Irvine, Birkebak.....5–10
Analysis for Aircraft Generators, Reliability. Duane, Yeager.....426–34
Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances. Lind, Schmitz.....476–9; disc. 479
Analysis of Nonlinear Sampled-Data Control Systems. Kinnen, Tou:
Part I.....386–90; disc. 394
Part II.....390–4; disc. 394
Anode Breaker Testing with a High-Capacity Rectifier System. Harris, Scully, Stewart.....43–9
Application of Continuous System Design Concepts to the Design of Sampled Data Systems. Schmidt.....74–8; disc. 78
Application of Switching Transistors and Saturable Reactors in a High-Performance Servo. Cox, Johannessen.....466–74
Arc Silicon-Smelting Furnace, Load Characteristics of a Submerged-. Grant.....273–5; disc. 275
Automatic Gage Control, The Use of Frequency Response Tests in the Analysis of a Foil Mill. Jones, Sills.....59–67
Automatic Gage Controller for a 56-Inch Reversing Steel Mill, An. Duke, Hulls.....10–15; disc. 15

B

Battery Impedance: Farads, Millionhms, Microhenrys. Willihngans, Rohner.....259–62
Bed Design for Cathodic Protection of Underground Structures, Trends in Ground. Landry, Howell.....456–61; disc. 461
Beds for Cathodic Protection, Deep Ground. Trouard, Wagner.....278–85
Boolean Functions, New Methods of Simplifying. Howard.....134–42; disc. 142
Breaker Testing with a High-Capacity Rectifier System, Anode. Harris, Scully, Stewart.....43–8
Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading, Notes on. Chandaket, Rosenstein.....148–58; disc. 158
Brushes, The Effect of Variable High-Altitude Humidity on the Wear of Nondusting. Moberly, Johnson..263–7
Bus for Shipboard Application, Aluminum. Thompson, Behr.....239–47; disc. 247

C

Car History and New Developments, Detector. Keevil.....233–8; disc. 238
Carrier Signal Generator for Servosystem Instrumentation, A Suppressed-. Pollard.....481–5
Cathodic Protection Applications at the Hanford Works. Bucholz.....394–9
Cathodic Protection, Deep Ground Beds for. Trouard, Wagner.....278–85
Cathodic Protection of Underground Structures, Trends in Ground Bed Design for. Landry, Howell.....456–61; disc. 461
Characteristics and Measurement of Ripple in Aircraft Electric Power Systems. Markowitz.....418–23; disc. 423
Characteristics of a Submerged-Arc Silicon-Smelting Furnace, Load. Grant.....273–5; disc. 275
Chart Relating Open-Loop and Closed-Loop Frequency Responses of Linear Control Systems, A New. Chen, Shen.....252–5
(Chemical Industry) Battery Impedance: Farads, Millionhms, Microhenrys. Willihngans, Rohner.259–62
(Chemical Industry) Cathodic Protection Applications at the Hanford Works. Bucholz.....394–9
(Chemical Industry) Deep Ground Beds for Cathodic Protection. Trouard, Wagner.....278–85
(Chemical Industry) High-Speed Restarting and Protection of Large Synchronous Motors. Phillips, Yuen.....360–7; disc. 367
(Chemical Industry) Trends in Ground Bed Design for Cathodic Protection of Underground Structures. Landry, Howell.....456–61; disc. 461
Chicago, Rapid Transit Is Expanding in. Forsythe.....474–6
Class A Random-Wound Motor Insulation as Determined by AIEE Standard No. 510 Test Procedure, The Life Expectancy of. (Committee Report).....224–8
Closed-Loop Frequency Responses of Linear Control Systems, A New Chart Relating Open-Loop and. Chen, Shen.....252–5
Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems. Elgerd, Stephens.....121–7
Coefficient of Performance, Significance of Heat Pump. Bary.....90–5; disc. 95
Coils, Three-Phase Induction Heating. Ross.....291–4
(Committee Report) The Life Expectancy of Class A Random-Wound Motor Insulation as Determined by AIEE Standard No. 510 Test Procedure. (58-1334).....224–8
Communication Systems for Railway Traffic Control. Sibley.....30–5
Compensation and 4-Terminal Network Loading, Notes on Bridged-T Complex Conjugate. Chandaket, Rosenstein.....148–58; disc. 158
Complex Conjugate Compensation and 4-Terminal Network, Loading, Notes on Bridged-T. Chandaket, Rosenstein.....148–58; disc. 158
Components, Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic. Sisco, Fiskin.....352–6
Components, The Electrical Analog for Determining Temperature Distribution in Electrical. Eckert, Hartnett, Irvine, Birkebak.....5–10
Computer Evaluation of High-Temperature Aircraft A-C Electrical System Designs. Sollecito, Swann.....434–44; disc. 444
Computer, Predicting the Performance of a Wind Tunnel Regulating System Using an Analog. Herbst, Nichols, Keay, Miller.....248–52
Considerations of an Ion Propulsion System, Practical. Lennert.....326–31; disc. 331
Considerations, Solar-Powered Thermoelectric Generator Design. Schuh, Tallent.....345–52
Constant Frequency A-C Power Using Variable Speed Generation. Jessee, Spaven....411–16; disc. 417
Constant-Frequency Devices: A Survey of the Methods in Use and Proposed, Variable-Speed. Owen.....321–5; disc. 325
Constant-Frequency Generating System, Energy-Conversion Properties of Induction Machines in Variable-Speed. Rias.....25–30
Constant-Frequency Generator System for Aircraft, Variable-Speed. Chirgwin, Stratton.....304–09; disc. 309
Constant-Frequency Variable-Speed Frequency-Make-up Generators. Hoard.....297–301; disc. 301
Control by Stochastic Adjustment. Bertram....485–91
Control Equipment, Functional Cycling to Assure Reliability of Aircraft. Hulsey, Kessler.....356–60
Control of Infrared Radiation. Barber.....54–9
Control of Multiactuator Systems, The Optimum. McCausland.....67–73
Control-System Performance, Probabilistic Error as a Measure of. Zaborszky, Diesel.....163–8
Control System, Signal Stabilization of a. Oldenburger, Liu.....96–100; disc. 100
Control Systems, A New Chart Relating Open-Loop and Closed-Loop Frequency Responses of Linear. Chen, Shen.....252–5
Control Systems, Analysis of Nonlinear Sampled-Data. Kinnen, Tou:
Part I.....386–90; disc. 394
Part II.....390–4; disc. 394
Control Systems, The Linear Least Squares Synthesis of Multivariable. Amara.....115–19; disc. 119
Control Systems, Use of a Mathematical Error Criterion in the Design of Adaptive. Merriam.....506–12
Controlled Adaptive Systems, Executive-. Staffin.....523–30
Controlled Reversible Induction-Motor Drives, Elements of Reactor-. Leonhard..106–14; disc. 114
Controller, A Practical Standard Transistorized Optimum Response. Chen, Little...337–45; disc. 345

Controllers for Rubber and Plastics Industries, Proposed Recommended Practices for Medium-Voltage Motor. Watkins.....504–06; disc. 506
Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems, Energy-. Rias.....25–30
Copper Joints, The Effect of Elevated Temperature on Flash-Welded Aluminum-. Dixon, Nelson.....491–5; disc. 495
Coulomb Friction and Stiction, Stability Criteria for Instrument Servomechanisms with. Pastel, Thaler..294–7
Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction, Stability. Pastel, Thaler..294–7
Criterion for Nonlinear Systems, A Stability. Ku, Wolf.....144–7; disc. 147
Criterion in the Design of Adaptive Control Systems, Use of a Mathematical Error. Merriam.....506–12
Cycling to Assure Reliability of Aircraft Control Equipment, Functional. Hulsey, Kesler.....356–60

D

Data Systems, Application of Continuous System Design Concepts to the Design of Sampled. Schmidt.....74–8; disc. 78
D-C Electric Drilling Rigs: Application and Operation. Hogwood.....286–9; disc. 289
D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment, Improved. Odok, Soelaiman.....186–93; disc. 193
Dead Time and Sampling, Relay-Type Feedback Control Systems with. Izawa, Weaver. 49–53; disc. 53
Deep Ground Beds for Cathodic Protection. Trouard, Wagner.....278–85
Describing-Function Technique, Limit-Cycle Stability Study of a Feedback Control System by a New. Harrington.....129–33; disc. 134
Design Concepts to the Design of Sampled Data Systems, Application of Continuous System. Schmidt.....74–8; disc. 78
Design Considerations for Large Industrial Semiconductor Rectifiers. Bowar.....1–5
Design Considerations, Solar-Powered Thermoelectric Generator. Schuh, Tallent.....345–52
Design Discussion of Thermoelectric Generation, An Elementary. Bollmeier.....445–50
Designs of Electric Resistance Heating, Some Unusual. Hynes.....40–3
Detector Car History and New Developments. Keevil.....233–8; disc. 238
Development for Hypersonic Aircraft, Generator Insulation Systems. Penn, Balke, Precopio.....255–9
Development of Fuses and Terminals for High-Temperature Applications. Bonwitt, Buttner.....267–72
Devices: A Survey of the Methods in Use and Proposed, Variable-Speed Constant-Frequency. Owen.....321–5; disc. 325
Dielectric Materials at 915 Mc, Rapid Heating of. Feiker, Gittinger.....35–9
Diesel and Electric Locomotives on the Milwaukee Road, Multiple-Unit Operation of. Wylie.....316–20; disc. 320
Discussion of Thermoelectric Generation, An Elementary Design. Bollmeier.....445–50
(Domestic and Commercial) Electrified Fibrous Air Filters. Thomas, Woodfin.....276–8
Drilling Rigs: Application and Operation, D-C Electric. Hogwood.....286–9; disc. 289
Drive at Southland Paper Mills, Transistor Amplistat-Regulated Sectional. Vickery, Shaad.....369–74; disc. 374
Drives, Elements of Reactor-Controlled Reversible Induction-Motor. Leonhard.....106–14; disc. 114
Dual-Mode Relay Servos. Buland, Furumoto.....405–11

E

Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems. Elgerd, Stephens.....121–7
Effect of Elevated Temperature on Flash-Welded Aluminum-Copper Joints, The. Dixon, Nelson.....491–5; disc. 495
Effect of Variable High-Altitude Humidity on the Wear of Nondusting Brushes, The. Moberly, Johnson.....263–7
Electric and Magnetic Conditions Inside an Inductively Heated Workpiece. Tudbury.....79–82; disc. 83
Electric Drilling Rigs: Application and Operation, D-C. Hogwood.....286–9; disc. 289
(Electric Heating) Control of Infrared Radiation. Barber.....54–9
Electric Locomotives on the Milwaukee Road, Multiple-Unit Operation of Diesel and. Wylie.....316–20; disc. 320
Electric Power Systems, Characteristics and Measurement of Ripple in Aircraft. Markowitz.....418–23; disc. 423
Electric Resistance Heating, Some Unusual Designs of. Hynes.....40–3
Electrical Analog for Determining Temperature Distribution in Electrical Components. Eckert, Hartnett, Irvine, Birkebak.....5–10
Electrical and Electronic Components, Some Effects of Hypersonic Ionization on the Design of. Sisco, Fiskin.....352–6
Electrical Features of the Four Corners Pipe Line. Sonnier, Siler.....209–23
Electrical System Designs, Computer Evaluation of High-Temperature Aircraft A-C. Sollecito, Swann.....434–44; disc. 444
Electrical Systems, A Solid Rotor A-C Generator for High-Temperature. Bateman.....400–03; disc. 403
Electrified Fibrous Air Filters. Thomas, Woodfin.....276–8
Electronic Components, Some Effects of Hypersonic Ionization on the Design of Electrical and. Sisco, Fiskin.....352–6
Electronic Instruments, Nuclear Radiation and. Crittenden.....423–6
(Electronic) Transient Voltages in Rectifier Transformers. Biega, Lord.....83–9; disc. 89
Elementary Design Discussion of Thermoelectric Generation, An. Bollmeier.....445–50
Elements of Reactor-Controlled Reversible Induction-Motor Drives. Leonhard.....106–14; disc. 114
Elevated Temperature on Flash-Welded Aluminum-Copper Joints, The Effect of. Dixon, Nelson.....491–5; disc. 495
Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems. Rias.....25–30
Equipment, Functional Cycling to Assure Reliability of Aircraft Control. Hulsey, Kesler.....356–60
Equipment, Improved D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C. Odok, Soelaiman.....186–93; disc. 193
Equipment, Temperature Prediction in Thermal-Lag. Richards.....462–5
Errratum.....330
Error as a Measure of Control-System Performance, Probabilistic. Zaborszky, Diesel.....163–8
Error Criterion in the Design of Adaptive Control Systems, Use of a Mathematical. Merriam.....506–12
Evaluation of High-Temperature Aircraft A-C Electrical System Designs, Computer. Sollecito, Swann.....434–44; disc. 444
Evaluation of Transient System Response. De Mello.....177–86
Executive-Controlled Adaptive Systems. Staffin.....523–30

F

Farads, Milliohms, Microhenrys, Battery Impedance: Willihnganz, Rohner.....259–62
Feedback Amplifiers, Transient Response and the Stabilization of. Mulligan.....495–503
(Feedback Control) A New Chart Relating Open-Loop and Closed-Loop Frequency Responses of Linear Control Systems. Chen, Shen.....252–5
(Feedback Control) A Practical Standard Transistorized Optimum Response Controller. Chien, Little.....337–45; disc. 345
(Feedback Control) A Stability Criterion for Nonlinear Systems. Ku, Wolf.....144–7; disc. 147
(Feedback Control) A Study of Nonlinear Systems with Random Inputs. Chuang, Kasda.....100–05
(Feedback Control) A Suppressed-Carrier Signal Generator for Servosystem Instrumentation. Pollard.....481–5
(Feedback Control) Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances. Lind, Schmitz.....476–9; disc. 479
(Feedback Control) Analysis of Nonlinear Sampled-Data Control Systems. Kitnen, Tou:
Part I.....386–90; disc. 394
Part II.....390–4; disc. 394
(Feedback Control) Application of Continuous System Design Concepts to the Design of Sampled Data Systems. Schmidt.....74–8; disc. 78
(Feedback Control) Application of Switching Transistors and Saturable Reactors in a High-Performance Servo. Cox, Johannessen.....466–74
(Feedback Control) Control by Stochastic Adjustment. Bertram.....485–91
(Feedback Control) Dual-Mode Relay Servos. Buland, Furumoto.....405–11
(Feedback Control) Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems. Elgerd, Stephens.....121–7
(Feedback Control) Evaluation of Transient System Response. De Mello.....177–86
(Feedback Control) Executive-Controlled Adaptive Systems. Staffin.....523–30
(Feedback Control) Input-Output Relationships for Multisampled-Loop Systems. Lendaris, Jury.....375–82; disc. 382
(Feedback Control) Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading. Chandaket, Rosenstein.....148–58; disc. 158
(Feedback Control) Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer. Herbst, Nichols, Keay, Miller.. 248–52
(Feedback Control) Probabilistic Error as a Measure of Control-System Performance. Zaborszky, Diesel.....163–8
(Feedback Control) Signal Stabilization of a Control System. Oldenburger, Liu.....96–100; disc. 100
(Feedback Control) Stability Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction. Pastel, Thaler.....294–7
Feedback Control System by a New Describing-Function Technique, Limit-Cycle Stability Study of a. Harrington.....129–33; disc. 134
Feedback Control System Design for Random Inputs, A Relay-Type. Hopkin, Wang.....228–33
Feedback Control Systems with Dead Time and Sampling, Relay-Type. Izawa, Weaver.....49–53; disc. 53
(Feedback Control) The Linear Least Squares Synthesis of Multivariable Control Systems. Amara.....115–19; disc. 119
(Feedback Control) The Optimum Control of Multiactuator Systems. McCausland.....67–73
(Feedback Control) Transfer Functions of Loaded Synchronous Machine. Hamdi-Sepen.....19–24
Feedback Systems, A Method for the Symbolic Representation and Analysis of Linear Periodic. Gilbert.....512–22; disc. 522
Fibrous Air Filters, Electrified. Thomas, Woodfin.....276–8
56-Inch Reversing Steel Mill, An Automatic Gage Controller for a. Duke, Hulls.....10–15; disc. 15
Filters, Electrified Fibrous Air. Thomas, Woodfin.....276–8
Flash-Welded Aluminum-Copper Joints, The Effect of Elevated Temperature on. Dixon, Nelson.....491–5; disc. 495
Foil Mill Automatic Gage Control, The Use of Frequency Response Tests in the Analysis of a. Jones, Sills.....59–67
Four Corners Pipe Line, Electrical Features of the. Sonnier, Siler.....209–23
4-Terminal Network Loading, Notes on Bridged-T Complex Conjugate Compensation and. Chandaket, Rosenstein.....148–58; disc. 158
Frequency A-C Power Using Variable Speed Generation, Constant. Jessee, Spaven.....411–16; disc. 417
Frequency Devices: A Survey of the Methods in Use and Proposed, Variable-Speed Constant-. Owen.....321–5; disc. 325
Frequency Generator System for Aircraft, Variable-Speed Constant-. Chirgwin, Stratton.....304–09; disc. 309
Frequency Response Tests in the Analysis of a Foil Mill Automatic Gage Control, The Use of. Jones, Sills.....59–67
Frequency Responses of Linear Control Systems, A New Chart Relating Open-Loop and Closed-Loop. Chen, Shen.....252–5
Frequency Variable-Speed Frequency-Make-up Generators, Constant-. Hoard.....297–301; disc. 301
Friction and Stiction, Stability Criteria for Instrument Servomechanisms with Coulomb. Pastel, Thaler.....294–7
Function Pole and Zero Locations on the Transient Response of Linear Control Systems, Effect of Closed-Loop Transfer. Elgerd, Stephens.....121–7
Functional Cycling to Assure Reliability of Aircraft Control Equipment. Hulsey, Kesler.....356–60
Functions, New Methods of Simplifying Boolean. Howard.....134–42; disc. 142
Functions of Loaded Synchronous Machine, Transfer. Hamdi-Sepen.....19–24
Furnace, Load Characteristics of a Submerged-Arc Silicon-Smelting. Grant.....275–5; disc. 275
Fuses and Terminals for High-Temperature Applications, Development of. Bonwitt, Buttner.....267–72

G

Gage Control, The Use of Frequency Response Tests in the Analysis of a Foil Mill Automatic. Jones, Sills.....59–67
Gage Controller for a 56-Inch Reversing Steel Mill, An Automatic. Duke, Hulls.....10–15; disc. 15
Generating System for Motor Coasters, A High-Capacity Maintenance-Free. Larson.....311–16
Generating Systems, Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency. Rias.....25–30
Generation, An Elementary Design Discussion of Thermoelectric. Bollmeier.....445–50
Generation, Constant Frequency A-C Power Using Variable Speed. Jessee, Spaven.....411–16; disc. 417
Generator Design Considerations, Solar-Powered Thermoelectric. Schuh, Tallent.....345–52

Generator for High-Temperature Electrical Systems, A Solid Rotor A-C. Bateman....400–03; disc. 403
Generator for Servosystem Instrumentation, A Suppressed-Carrier Signal. Pollard....481–5
Generator Insulation Systems Development for Hypersonic Aircraft. Penn, Balke, Precopio....255–9
Generator, Optimum Reflector-Absorber Geometry for a Solar. Sitneman....332–7
Generator System for Aircraft, Variable-Speed Constant-Frequency. Chirgwin, Stratton.304–09; disc. 309
Generators, Constant-Frequency Variable-Speed Frequency-Make-up. Hoard....297–301; disc. 301
Generators, Reliability Analysis for Aircraft. Duane, Yeager....426–34
Geometry for a Solar Generator, Optimum Reflector-Absorber. Sitneman....332–7
Ground Bed Design for Cathodic Protection of Underground Structures, Trends in. Landry, Howell....456–61; disc. 461
Ground Beds for Cathodic Protection, Deep. Trouard, Wagner....278–85

H

Hanford Works, Cathodic Protection Applications at the. Bucholz....394–9
Heat Pump Coefficient of Performance, Significance of. Bary....90–5; disc. 95
Heated Workpiece, Electric and Magnetic Conditions Inside an Induction-. Tudbury....79–82; disc. 83
Heating Coils, Three-Phase Induction. Ross....291–4
Heating of Dielectric Materials at 915 Mc, Rapid. Feiker, Gittinger....35–9
Heating, Some Unusual Designs of Electric Resistance. Hynes....40–3
High-Altitude and Ultrahigh Temperature Electronic Transformer Applications, Terminal Spacings for. Duncan, Rectanus....16–19
High-Altitude Humidity on the Wear of Nondusting Brushes, The Effect of Variable. Moberly, Johnson....263–7
High-Capacity Maintenance-Free Generating System for Motor Coaches, A. Larson....311–16
High-Capacity Rectifier System, Anode Breaker Testing with a. Harris, Scully, Stewart....43–8
High-Performance Servo, Application of Switching Transistors and Saturable Reactors in a. Cox, Johannesen....466–74
High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment, Improved D-C. Odok, Soelaiman....186–93; disc. 193
High-Speed Restarting and Protection of Large Synchronous Motors. Phillips, Yuen....360–7; disc. 367
High-Temperature Aircraft A-C Electrical System Design, Computer Evaluation of. Sifleet, Swan....434–44; disc. 444
High-Temperature Aircraft, New Materials for Transformer-Rectifier Unit in. Hoop, McIlvaine....450–6
High-Temperature Applications, Development of Fuses and Terminalsfor. Bonwit, Buttner....267–72
High-Temperature Electrical Systems, A Solid Rotor A-C Generator for. Bateman....400–03; disc. 403
History and New Developments, Detector Car. Keevil....233–8; disc. 238
Humidity on the Wear of Nondusting Brushes, The Effect of Variable High-Altitude. Moberly, Johnson....263–7
Hypersonic Aircraft, Generator Insulation Systems Developmentfor. Penn, Balke, Precopio....255–9
Hypersonic Ionization on the Design of Electrical and Electronic Components, Some Effects of. Sisco, Fiskin....352–6

I

Impedance: Farads, Milliohms, Microhenrys, Battery. Willihnganz, Rohner....259–62
Impedances, Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source. Lind, Schmitz....476–9; disc. 479
Improved D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment. Odok, Soelaiman....186–93; disc. 193
Induction-Heated Workpiece, Electric and Magnetic Conditions Inside an. Tudbury....79–82; disc. 83
Induction Heating Coils, Three-Phase. Ross....291–4
Induction Machines in Variable-Speed Constant-Frequency Generating Systems, Energy-Conversion Properties of. Rias....25–30
Induction-Motor Drives, Elements of Reactor-Controlled Reversible. Leonard....106–14; disc. 114
(Industrial and Commercial) Load Characteristics of a Submerged-Arc Silicon-Smelting Furnace. Grant....273–5; disc. 275
(Industrial Control) Elements of Reactor Controlled Reversible Induction-Motor Drives. Leonhard....106–14; disc. 114
(Industrial Control) New Methods of Simplifying Boolean Functions. Howard....134–42; disc. 142
Industrial Semiconductor Rectifiers, Design Considerations for Large. Bowar....1–5
Infrared Radiation, Control of. Barber....54–9
Input-Output Relationships for Multisampled-Loop Systems. Lendaris, Jury....374–82; disc. 382
Inputs, A Relay-Type Feedback Control System Design for Random. Hopkin, Wang....228–33
Inputs, A Study of Nonlinear Systems with Random. Chuang, Kazda....100–05
Instrument Servomechanisms with Coulomb Friction and Friction, Stability Criteria for. Pastel, Thaler....294–7
Instrumentation, A Recommended Program for Resistance-Welding. Dixon....199–209
Instrumentation, A Suppressed-Carrier Signal Generator for Servosystem. Pollard....481–5
Instruments, Nuclear Radiation and Electronic. Critenden....423–6
Insulation as Determined by AIEE Standard No. 510 Test Procedure, The Life Expectancy of Class A Random-Wound Motor. (Committee Report)....224–8
Insulation Systems Development for Hypersonic Aircraft, Generator. Penn, Balke, Precopio....255–9
Insulation Systems in Low- and Medium-Voltage D-C Equipment, Improved D-C High-Potential Testing of. Odok, Soelaiman....186–93; disc. 193
Ion Propulsion System, Practical Considerations of an. Lennert....326–31; disc. 331
Ionization on the Design of Electrical and Electronic Components, Some Effects of Hypersonic. Sisco, Fiskin....352–6

J

Joints, The Effect of Elevated Temperature on Flash-Welded Aluminum-Copper. Dixon, Nelson....491–5; disc. 495

L

(Land Transportation) A High-Capacity Maintenance-Free Generating System for Motor Coaches. Larson....311–16
(Land Transportation) Communication Systems for Railway Traffic Control. Sibley....30–5
(Land Transportation) Detector Car History and New Developments. Keevil....233–8; disc. 238
(Land Transportation) Improved D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment. Odok, Soelaiman....186–93; disc. 193
(Land Transportation) Modern Systems of Traffic Control as Applied to the Seaboard Air Line Railroad Company. DePriest....168–72
(Land Transportation) Multiple-Unit Operation of Diesel and Electric Locomotives on the Milwaukee Road. Watkins....316–20; disc. 320
(Land Transportation) Rapid Transit Is Expanding in Chicago. Forsythe....474–6
(Land Transportation) Traffic Control for Railroads. Baughman....173–6
Life Expectancy of Class A Random-Wound Motor Insulation as Determined by AIEE Standard No. 510 Test Procedure, The. (Committee Report)....224–8
Limit-Cycle Stability Study of a Feedback Control System by a New Describing-Function Technique. Harrington....129–33; disc. 134
Linear Control Systems, A New Chart Relating Open-Loop and Closed-Loop Frequency Responses of. Chen, Shen....252–5
Linear Control Systems, Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of. Elgerd, Stephens....121–7
Linear Least Squares Synthesis of Multivariable Control Systems, The. Amara....115–19; disc. 119
Linear Periodic Closed-Loop Systems, A Method for the Symbolic Representation and Analysis of. Gilbert....512–22; disc. 522
Load Characteristics of a Submerged-Arc Silicon-Smelting Furnace. Grant....273–5; disc. 275
Loaded Synchronous Machine, Transfer Functions of. Hamdi-Sepen....9–24
Loading, Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network. Chandaket, Rosenstein....148–58; disc. 158
Locomotives on the Milwaukee Road, Multiple-Unit Operation of Diesel and Electric. Wylie....316–20; disc. 320
Loop Systems, Input-Output Relationships for Multisampled-. Lendaris, Jury....375–82; disc. 382
Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems, Effect of Closed-. Elgerd, Stephens....121–7
Low- and Medium-Voltage D-C Equipment, Improved D-C High-Potential Testing of Insulation Systems in. Odok, Soelaiman....186–93; disc. 193

M

Machines in Variable-Speed Constant-Frequency Generating Systems, Energy-Conversion Properties of Induction. Rias....25–30
Magnetic Conditions Inside an Induction-Heated Workpiece, Electric and. Tudbury....79–82; disc. 83
Maintenance-Free Generating System for Motor Coaches, A High-Capacity. Larson....311–16
(Marine Transportation) Aluminum Bus for Shipboard Application. Thompson, Behr....239–47; disc. 247
Materials for Transformer-Rectifier Unit in High-Temperature Aircraft, New. Hoop, McIlvaine....450–6
Mathematical Error Criterion in the Design of Adaptive Control Systems, Use of a. Merriam....506–12
Measure of Control-System Performance, Probabilistic Error as a. Zaborszky, Diesel....163–8
Measurement of Ripple in Aircraft Electric Power Systems, Characteristics and. Markowitz....418–23; disc. 423
(Metal Industry) An Automatic Gage Controller for a 56-Inch Reversing Steel Mill. Duke, Hulls....10–15; disc. 15
(Metal Industry) The Use of Frequency Response Tests in the Analysis of a Foil Mill Automatic Gage Control. Jones, Sills....59–67
Method for the Symbolic Representation and Analysis of Linear Periodic Feedback Systems, A. Gilbert....512–22; disc. 522
Methods of Simplifying Boolean Functions, New. Howard....134–42; disc. 142
Microhenrys, Battery Impedance: Farads, Milliohms, Willihnganz, Rohner....259–62
Mill, An Automatic Gage Controller for a 56-Inch Reversing Steel. Duke, Hulls....10–15; disc. 15
Mill Automatic Gage Control, The Use of Frequency Response Tests in the Analysis of a Foil. Jones, Sills....59–67
Milliohms, Microhenrys, Battery Impedance: Farads, Willihnganz, Rohner....259–62
Milwaukee Road, Multiple-Unit Operation of Diesel and Electric Locomotives on the. Wylie....316–20; disc. 320
Mode Relay Servos, Dual-. Buland, Furumoto....405–11
Modern Systems of Traffic Control as Applied to the Seaboard Air Line Railroad Company. DePriest....168–72
Motor Coaches, A High-Capacity Maintenance-Free Generating System for. Larson....311–16
Motor Controllers for Rubber and Plastics Industries, Proposed Recommended Practices for Medium-Voltage. Watkins....504–06; disc. 506
Motor Drives, Elements of Reactor-Controlled Reversible Induction-. Leonhard....106–14; disc. 114
Motor Insulation as Determined by AIEE Standard No. 510 Test Procedure, The Life Expectancy of Class A Random-Wound. (Committee Report)....224–8
Motors, High-Speed Restarting and Protection of Large Synchronous. Phillips, Yuen....360–7; disc. 367
Multiactuator Systems, The Optimum Control of. McCausland....67–73
Multisampled-Loop Systems, Input-Output Relationships for. Lendaris, Jury....375–82; disc. 382
Multivariable Control Systems, The Linear Least Squares Synthesis of. Amara....115–19; disc. 119

N

Network Loading, Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal. Chandaket, Rosenstein....148–58; disc. 158
New Methods of Simplifying Boolean Functions. Howard....134–42; disc. 142
915 Mc, Rapid Heating of Dielectric Materials at. Feiker, Gittinger....35–9
Nonlinear Discontinuous Source Impedances, Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and. Lind, Schmitz....476–9; disc. 479
Nonlinear Sampled-Data Control Systems, Analysis of. Kinnen, Tou:
Part I....386–90; disc. 394
Part II....390–4; disc. 394
Nonlinear Systems, A Stability Criterion for. Ku, Wolf....144–7; disc. 147
Nonlinear Systems with Random Inputs, A Study of. Chuang, Kazda....100–05
Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading. Chandaket, Rosenstein....148–58; disc. 158
Nuclear Radiation and Electronic Instruments. Critenden....423–6

O

Open-Loop and Closed-Loop Frequency Responses of Linear Control Systems, A New Chart Relating. Chen, Shen....252–5

Optimum Control of Multiactuator Systems, The. McCausland.....67–73
Optimum Reflector-Absorber Geometry for a Solar Generator. Stineman.....332–7
Optimum Response Controller, A Practical Standard Transistorized. Chen, Little.....337–45; disc. 345
Output Relationships for Multisampled Loop Systems, Input-. Lendaris, Jury.....375–82; disc. 382

P

Paper Mills, Transistor Amplistat-Regulated Sectional Drive at Southland. Vickery, Shaad.....369–74; disc. 374
Performance, Probabilistic Error as a Measure of Control-System. Zaborszky, Diesel.....163–8
Performance, Significance of Heat Pump Coefficient of. Bary.....90–5; disc. 95
(Petroleum Industry) D-C Electric Drilling Rigs: Application and Operation. Hogwood.....286–9; disc. 289
(Petroleum Industry) Electrical Features of the Four Corners Pipe Line. Sonnier, Siler.....209–23
Pipe Line, Electrical Features of the Four Corners. Sonnier, Siler.....209–23
Pole and Zero Locations on the Transient Response of Linear Control Systems, Effect of Closed-Loop Transfer Function. Elgerd, Stephens.....121–7
Power Systems, Characteristics and Measurement of Ripple in Aircraft Electric. Markowitz.....418–23; disc. 423
Power Using Variable Speed Generation, Constant Frequency A-C. Jesser, Spaven.. 411–16; disc. 417
Practical Considerations of an Ion Propulsion System. Lennert.....326–31; disc. 331
Practical Standard Transistorized Optimum Response Controller, A. Chen, Little.....337–45; disc. 345
Practices for Medium-Voltage Motor Controllers for Rubber and Plastics Industries, Proposed Recommended. Watkins.....504–06; disc. 506
Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer. Herbst, Nichols, Keay, Miller.....248–52
Probabilistic Error as a Measure of Control-System Performance. Zaborszky, Diesel.....163–8
Program for Resistance-Welding Instrumentation, A Recommended. Dixon.....199–209
Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems, Energy-Conversion. Riaz.....25–30
Proposed Recommended Practices for Medium-Voltage Motor Controllers for Rubber and Plastics Industries. Watkins.....504–06; disc. 506
Propulsion System, Practical Considerations of an Ion. Lennert.....326–31; disc. 331
Protection of Large Synchronous Motors, High-Speed Restarting and. Phillips, Yuen.. 360–7; disc. 367
Pump Coefficient of Performance, Significance of Heat. Bary.....90–5; disc. 95

R

Radiation and Electronic Instruments, Nuclear. Crittenden.....423–6
Radiation, Control of Infrared. Barber.....54–9
Railroads, Traffic Control for. Baughman.....173–6
Railway Traffic Control, Communication Systems for. Sibley.....30–5
Random Inputs, A Relay-Type Feedback Control System Design for. Hopkin, Wang.....228–33
Random Inputs, A Study of Nonlinear Systems with. Chuang, Kazda.....100–05
Random-Wound Motor Insulation as Determined by AIEE Standard No. 510 Test Procedure, The Life Expectancy of Class A. (Committee Report).....224–8
Rapid Heating of Dielectric Materials at 915 Mc. Feiker, Gittinger.....35–9
Rapid Transit is Expanding in Chicago. Forsythe.....474–6
Reactor-Controlled Reversible Induction-Motor Drives, Elements of. Leonhard.....106–14; disc. 114
Recommended Practices for Medium-Voltage Motor Controllers for Rubber and Plastics Industries, Proposed. Watkins.....504–06; disc. 506
Recommended Program for Resistance-Welding Instrumentation, A. Dixon.....199–209
Rectifier System, Anode Breaker Testing with a High-Capacity. Harris, Scully, Stewart.....43–8
Rectifier Transformers, Transient Voltages in. Biega, Lord.....89–9; disc. 89
Rectifier Unit in High-Temperature Aircraft, New Materials for Transformer-. Hoop, McIlvaine.....450–6
Rectifiers, Design Considerations for Large Industrial Semiconductor. Bowar.....1–5
Reflector-Absorber Geometry for a Solar Generator, Optimum. Stineman.....332–7
Regulating System Using an Analog Computer, Predicting the Performance of a Wind Tunnel. Herbst, Nichols, Keay, Miller.....248–52
Relationships for Multisampled-Loop Systems, Input-Output. Lendaris, Jury.....375–82; disc. 382
Relay Servos, Dual-Mode. Buland, Furumoto.. 405–11
Relay-Type Feedback Control System Design for Random Inputs, A. Hopkin, Wang.....228–33
Relay-Type Feedback Control Systems with Dead Time and Sampling. Izawa, Weaver.....49–53; disc. 53
Reliability Analysis for Aircraft Generators. Duane, Yeager.....426–34
Representation and Analysis of Linear Periodic Feedback Systems, A Method for the Symbolic. Gilbert.....512–22; disc. 522
Resistance Heating, Some Unusual Designs of Electric. Hynes.....40–3
Resistance-Welding Instrumentation, A Recommended Program for. Dixon.....199–209
Response and the Stabilization of Feedback Amplifiers, Transient. Mulligan.....495–503
Response Controller, A Practical Standard Transistorized Optimum. Chen, Little.....337–45; disc. 345
Response, Evaluation of Transient System. DeMello.....177–86
Restarting and Protection of Large Synchronous Motors, High-Speed. Phillips, Yuen.. 360–7; disc. 367
Reversible Induction-Motor Drives, Elements of Reactor-Controlled. Leonhard.. 106–14; disc. 114
Reversing Steel Mill, An Automatic Gage Controller for a 56-Inch. Duke, Hulls.....10–15; disc. 15
Rigs: Application and Operation, D-C Electric Drilling. Hogwood.....286–9; disc. 289
Ripple in Aircraft Electric Power Systems, Characteristics and Measurement of. Markowitz.....418–23; disc. 423
(Rotating Machinery) Constant Frequency A-C Power Using Variable Speed Generation. Jesser, Spaven.....411–16; disc. 417
Rotor A-C Generator for High-Temperature Electrical Systems, A Solid. Bateman.....400–03; disc. 403
Rubber and Plastics Industries, Proposed Recommended Practices for Medium-Voltage Motor Controllers for. Watkins.....504–06; disc. 506

S

Sampled-Data Control Systems, Analysis of Nonlinear. Kinnen, Tou:
Part I.....386–90; disc. 394
Part II.....390–4; disc. 394
Sampled Data Systems, Application of Continuous System Design Concepts to the Design of. Schmidt.....74–8; disc. 78
Sampling, Relay-Type Feedback Control Systems with Dead Time and. Izawa, Weaver.. 49–53; disc. 53
Saturable Reactors in a High-Performance Servo, Application of Switching Transistors and. Cox, Johannesen.....466–74
Seaboard Air Line Railroad Company, Modern Systems of Traffic Control as Applied to the. De-Priest.....168–72
Sectional Drive at Southland Paper Mills, Transistor Amplistat-Regulated. Vickery, Shaad.....369–74; disc. 374
Semiconductor Rectifiers, Design Considerations for Large Industrial. Bowar.....1–5
Servo, Application of Switching Transistors and Saturable Reactors in a High-Performance. Cox, Johannesen.....466–74
Servomechanisms with Coulomb Friction and Stiction, Stability Criteria for Instrument. Pastel, Thaler.....294–7
Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances, Analysis of A-C. Lind, Schmitz.....476–9; disc. 479
Servos, Dual-Mode Relay. Buland, Furumoto.. 405–11
Servosystem Instrumentation, A Suppressed-Carrier Signal Generator for. Pollard.....481–5
Shipboard Application, Aluminum Bus for. Thompson, Behr.....239–47; disc. 247
Signal Generator Servosystem Instrumentation, A Suppressed-Carrier. Pollard.....481–5
Signal Stabilization of a Control System. Oldenburger, Liu.....96–100; disc. 100
Significance of Heat Pump Coefficient of Performance. Bary.....90–5; disc. 95
Silicon-Smelting Furnace, Load Characteristics of a Submerged-Arc. Grant.....273–5; disc. 275
Simplifying Boolean Functions, New Methods of. Howard.....134–42; disc. 142
Smelting Furnace, Load Characteristics of a Submerged-Arc Silicon-. Grant.....273–5; disc. 275
Solar Generator, Optimum Reflector-Absorber Geometry for a. Stineman.....332–7
Solar-Powered Thermoelectric Generator Design Considerations. Schuh, Tallent.....345–52
Solid Rotor A-C Generator for High-Temperature Electrical Systems, A. Bateman.....400–03; disc. 403
Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components. Sisco, Fiskin.....352–6
Some Unusual Designs of Electric Resistance Heating. Hynes.....40–3
Sources and Nonlinear Discontinuous Source Impedances, Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage. Lind, Schmitz.....476–9; disc. 479
Southland Paper Mills, Transistor Amplistat-Regulated Sectional Drive at. Vickery, Shaad.....369–74; disc. 374
Speed Generation, Constant Frequency A-C Power Using Variable. Jesser, Spaven.. 411–16; disc. 417
Square Synthesis of Multivariable Control Systems, The Linear Least. Amara.....115–19; disc. 119
Stability Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction. Pastel, Thaler.....294–7
Stability Criterion for Nonlinear Systems, A. Ku, Wolf.....144–7; disc. 147
Stability Study of a Feedback Control System by a New Describing-Function Technique, Limit-Cycle. Harrington.....129–33; disc. 134
Stabilization of a Control System, Signal. Oldenburger, Liu.....96–100; disc. 100
Stabilization of Feedback Amplifiers, Transient Response and the. Mulligan.....495–503
Standard No. 510 Test Procedure, The Life Expectancy of Class A Random-Wound Motor Insulation as Determined by AIEE. (Committee Report).....224–8
Standard Transistorized Optimum Response Controller, A Practical. Chen, Little.....337–45; disc. 345
Steel Mill, An Automatic Gage Controller for a 56-Inch Reversing. Duke, Hulls.....10–15; disc. 15
Stiction, Stability Criteria for Instrument Servomechanisms with Coulomb Friction and. Pastel, Thaler.....294–7
Stochastic Adjustment, Control by. Bertram.....485–91
Structures, Trends in Ground Bed Design for Cathodic Protection of Underground. Landry, Howell.....456–61; disc. 461
Study of a Feedback Control System by a New Describing-Function Technique, Limit-Cycle Stability. Harrington.....129–33; disc. 134
Study of Nonlinear Systems With Random Inputs, A. Chuang, Kazda.....100–05
Submerged-Arc Silicon-Smelting Furnace, Load Characteristics of a. Grant.....273–5; disc. 275
Suppressed-Carrier Signal Generator for Servosystem Instrumentation, A. Pollard.....481–5
Survey of the Methods in Use and Proposed, Variable-Speed Constant-Frequency Devices: A. Owen.....321–5; disc. 325
Switching Transistors and Saturable Reactors in a High-Performance Servo, Application of. Cox, Johannesen.....466–74
Symbolic Representation and Analysis of Linear Periodic Feedback Systems, A Method for the. Gilbert.....512–22; disc. 522
Synchronous Machine, Transfer Functions of Loaded. Hamdi-Sepen.....19–24
Synchronous Motors, High-Speed Restarting and Protection of Large. Phillips, Yuen.....360–7; disc. 367
Synthesis of Multivariable Control Systems, The Linear Least Squares. Amara.....115–19; disc. 119
System Design Concepts to the Design of Sampled Data Systems, Application of Continuous. Schmidt.....74–8; disc. 78
System Designs, Computer Evaluation of High-Temperature Aircraft A-C Electrical. Sollecito, Swann.....434–44; disc. 444
System Response, Evaluation of Transient. De Mello.....177–86
Systems, A Stability Criterion for Nonlinear. Ku, Wolf.....144–7; disc. 147
Systems: Development for Hypersonic Aircraft, Generator Insulation. Penn, Balke, Precopio.. 255–9
Systems, Executive-Controlled Adaptive. Staffin.....523–30
Systems, The Optimum Control of Multiactuator Systems. McCausland.....67–73
Systems with Random Inputs, A Study of Nonlinear. Chuang, Kazda.....100–05

T

Technique, Limit-Cycle Stability Study of a Feedback Control System by a New Describing-Function. Harrington.....129–33; disc. 134
Temperature Aircraft A-C Electrical System Designs, Computer Evaluation of High-. Sollecito, Swann.....434–44; disc. 444
Temperature Aircraft, New Materials for Transformer-Rectifier Units in High-. Hoop, McIlvaine.. 450–6
Temperature Applications, Development of Fuses and Terminals for High-. Bonwitt, Buttner.....267–72
Temperature Distribution in Electrical Components, The Electrical Analog for Determining. Eckert, Hartnett, Irvine, Birkebak.....5–10
Temperature Electronic Transformer Applications, Terminal Spacings for High-Altitude and Ultra-high-. Duncan, Rectanus.....16–19

Temperature on Flash-Welded Aluminum-Copper Joints, The Effect of Elevated. Dixon, Nelson....491–5; disc. 495
Temperature Prediction in Thermal-Lag Equipment. Richards....462–5
Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications. Duncan, Rectanus....16–19
Terminals for High-Temperature Applications, Development of Fuses and. Bonwitt, Buttner..267–72
Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment, Improved D-C High-Potential. Oded, Soelaiman....186–93; disc. 193
Testing with a High-Capacity Rectifier System, Anode Breaker. Harris, Scully, Stewart....43–8
Tests in the Analysis of a Foil Mill Automatic Gage Control, The Use of Frequency Response. Jones, Sills....59–67
Thermal-Lag Equipment, Temperature Prediction in. Richards....462–5
Thermoelectric Generation, An Elementary Design Discussion of. Bollmeier....445–50
Thermoelectric Generator Design Considerations, Solar-Powered. Schuh, Tallent....345–52
Three-Phase Induction Heating Coils. Ross....291–4
Traffic Control as Applied to the Seaboard Air Line Railroad Company, Modern Systems of. DePriest....168–72
Traffic Control, Communication Systems for Railway. Sibley....30–5
Traffic Control for Railroads. Baughman....173–6
Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems, Effect of Closed-Loop. Elgerd, Stephens....121–7
Transfer Functions of Loaded Synchronous Machine. Hamdi-Sepen....19–24
Transformer Applications, Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic. Duncan, Rectanus....16–19
Transformer-Rectifier Unit in High-Temperature Aircraft, New Materials for. Hoop, McIlvaine....450–6
Transformers, Transient Voltages in Rectifier. Biega, Lord....83–9; disc. 89
Transient Response and the Stabilization of Feedback Amplifiers. Mulligan....495–503
Transient Response of Linear Control Systems, Effect of Closed-Loop Transfer Function Pole and Zero Locations on the. Elgerd, Stephens....121–7
Transient System Response, Evaluation of. De Mello....177–86
Transient Voltages in Rectifier Transformers. Biega, Lord....83–9; disc. 89
Transistor Amplistat-Regulated Sectional Drive at Southland Paper Mills. Vickery, Shaad....369–74; disc. 374
Transistorized Optimum Response Controller, A Practical Standard. Chen, Little....337–45; disc. 345
Transistors and Saturable Reactors in a High-Performance Servo, Application of Switching. Cox, Johannessen....466–74
Transit Is Expanding in Chicago, Rapid. Forsythe....474–6
(Transportation, Air) A Solid Rotor A-C Generator for High Temperature Electrical Systems. Bateman....400–03; disc. 403
(Transportation, Air) An Elementary Design Discussion of Thermoelectric Generation. Bollmeier....445–50
(Transportation, Air) Constant-Frequency Variable-Speed Frequency-Make-up Generators. Hoard....297–301; disc. 301
(Transportation, Air) Development of Fuses and Terminals for High-Temperature Applications. Bonwitt, Buttner....267–72
(Transportation, Air) Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems. Riaz....25–30
(Transportation, Air) Nuclear Radiation and Electronic Instruments. Crittenden....423–6
(Transportation, Air) Optimum Reflector-Absorber Geometry for a Solar Generator. Stineman....332–7
(Transportation, Air) Practical Considerations of an Ion Propulsion System. Lennert....326–31; disc. 331
(Transportation, Air) Solar-Powered Thermoelectric Generator Design Considerations. Schuh, Tallent....345–52
(Transportation, Air) Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components. Sisco, Fiskin....352–6
(Transportation, Air) Temperature Prediction in Thermal-Lag Equipment. Richards....462–5
(Transportation, Air) Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications. Duncan, Rectanus....16–19
(Transportation, Air) The Effect of Variable High-Altitude Humidity on the Wear of Nondusting Brushes. Moberly, Johnson....263–7
(Transportation, Air) The Electrical Analog for Determining Temperature Distribution in Electrical Components. Eckert, Hartnett, Irvine, Birkebak....5–10
(Transportation, Air) Variable-Speed Constant-Frequency Devices: A Survey of the Methods in Use and Proposed. Owen....321–5; disc. 325
(Transportation, Land) A High-Capacity Maintenance-Free Generating System for Motor Coaches. Larson....311–16
(Transportation, Land) Communication Systems for Railway Traffic Control. Sibley....30–5
(Transportation, Land) Detector Car History and New Developments. Keevil....233–8; disc. 238
(Transportation, Land) Modern Systems of Traffic Control as Applied to the Seaboard Air Line Railroad Company. DePriest....168–72
(Transportation, Land) Multiple-Unit Operation of Diesel and Electric Locomotives on the Milwaukee Road. Wylie....316–20; disc. 320
(Transportation, Land) Rapid Transit Is Expanding in Chicago. Forsythe....474–6
(Transportation, Land) Traffic Control for Railroads. Baughman....173–6
(Transportation, Marine) Aluminum Bus for Shipboard Application. Thompson, Behr....239–47; disc. 247
Trends in Ground Bed Design for Cathodic Protection of Underground Structures. Landry, Howell....456–61; disc. 461

U

Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances, Analysis of A-C Servomotors Operated from. Lind, Schmitz....476–9; disc. 479
Underground Structures, Trends in Ground Bed Design for Cathodic Protection of. Landry, Howell....456–61; disc. 461
Use of a Mathematical Error Criterion in the Design of Adaptive Control Systems. Merriam....506–12
Use of Frequency Response Tests in the Analysis of a Foil Mill Automatic Gage Control, The. Jones, Sills....59–67

V

Variable High-Altitude Humidity on the Wear of Nondusting Brushes, The Effect of. Moberly, Johnson....263–7
Variable-Speed Constant-Frequency Devices: A Survey of the Methods in Use and Proposed. Owen....321–5; disc. 325
Variable-Speed Constant-Frequency Generating Systems, Energy-Conversion Properties of Induction Machines in. Riaz....25–30
Variable-Speed Constant-Frequency Generator System for Aircraft. Chirgwin, Stratton....304–09; disc. 309
Variable-Speed Frequency-Make-up Generators, Constant-Frequency. Hoard....297–301; disc. 301
Variable Speed Generation, Constant Frequency A-C Power Using. Esser, Spaven....411–16; disc. 417
Voltage Sources and Nonlinear Discontinuous Source Impedances, Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal. Lind, Schmitz....476–9; disc. 479
Voltages in Rectifier Transformers, Transient. Biega, Lord....83–9; disc. 89

W

Welded Aluminum-Copper Joints, The Effect of Elevated Temperature on Flash-. Dixon, Nelson....491–5; disc. 495
Welding Instrumentation, A Recommended Program for Resistance-. Dixon....199–209
Wind Tunnel Regulating System Using an Analog Computer, Predicting the Performance of a. Herbst, Nichols, Keay, Miller....248–52
Workpiece, Electric and Magnetic Conditions Inside an Induction-Heated. Tudbury....79–82; disc. 83
Wound Motor Insulation as Determined by AIEE Standard No. 510 Test Procedure, The Life Expectancy of Class A Random-. (Committee Report)....224–8

Z

Zero Locations on the Transient Response of Linear Control Systems, Effect of Closed-Loop Transfer Function Pole and. Elgerd, Stephens....121–7

2. Author Index

Amara, R. C. The Linear Least Squares Synthesis of Multivariable Control Systems. (59–214)....115–19; disc. 120
Ash, R. B.; W. H. Kim, G. M. Kranc. Disc....383

B

Baker, R. M. Disc....83
Baike, R. L.; W. B. Penn, F. M. Precopio. Generator Insulation Systems Development for Hypersonic Aircraft. (59–864)....255–9
Barber, Ira J. Control of Infrared Radiation. (59–31)....54–9
Bary, Constantine W. Significance of Heat Pump Coefficient of Performance. (59–20)....90–5; disc. 96
Bateman, J. T. A Solid Rotor A-C Generator for High-Temperature Electrical Systems. (59–848)....400–03; disc. 405
Baughman, G. W. Traffic Control for Railroads. (59–252)....173–6
Behr, S. H.; G. J. Thompson. Aluminum Bus for Shipboard Application. (59–654)....239–47; disc. 247
Bertram, J. E. Control by Stochastic Adjustment. (59–1156)....485–91
Bhimani, B. V. Disc....194
Biega, B. C.; H. W. Lord. Transient Voltages in Rectifier Transformers. (59–149)....83–9; disc. 90
Birkebak, Richard; E. R. G. Eckert, J. P. Hartnett, T. F. Irvine. The Electrical Analog for Determining Temperature Distribution in Electrical Components. (58–920)....5–10
Bollmeier, E. W. An Elementary Design Discussion of Thermoelectric Generation. (59–867)....445–50
Bond, N. T. Disc....247
Bonwitt, W. F.; H. Buttner. Development of Fuses and Terminals for High-Temperature Applications. (59–771)....267–72
Bowar, Gerald J. Design Considerations for Large Industrial Semiconductor Rectifiers. (58–223)....1–5
Brighman, F. P. Disc....367
Bucholz, C. S. Cathodic Protection Applications at the Hanford Works. (59–821)....394–9
Buland, R. M.; N. Furumoto. Dual-Mode Relay Servos. (59–840)....405–11
Bultman, E. W.; C. W. Martin. Disc....195
Buttner, H.; W. F. Bonwitt. Development of Fuses and Terminals for High-Temperature Applications. (59–771)....267–72

C

Cameron, A. W. W. Disc....195
Campbell, S. J.; E. C. Fox. Disc....374
Candee, A. H. Disc....290
Chandaket, Prapat; A. B. Rosenstein. Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading. (59–220)....148–58; disc. 162
Chen, C. F.; D. W. C. Shen. A New Chart Relating Open-Loop and Closed-Loop Frequency Responses of Linear Control Systems. (59–647)....252–5
Chen, Kan; D. R. Little. A Practical Standard Transistorized Optimum Response Controller. (59–838)....337–45; disc. 345
Chestnut, Harold. Disc....345
Chirgwin, K. M. Disc....403
Chirgwin, K. M.; L. J. Stratton. Variable-Speed Constant-Frequency Generator System for Aircraft. (59–872)....304–09; disc. 309
Chirgwin, K. M.; L. F. Stratton. Disc....301, 417
Chuang, Kuei; L. F. Kazda. A Study of Nonlinear Systems with Random Inputs. (59–147)....100–05
Colehower, E. W. Disc....423
Correy, Thomas B. Disc....495
Cox, Fred B.; F. R. Johannessen. Application of Switching Transistors and Saturable Reactors in a High-Performance Servo. (58–89)....466–74
Crittenden, J. R. Nuclear Radiation and Electronic Instruments. (59–873)....423–6
Curdts, E. B. Disc....195

D

De Mello, F. P. Evaluation of Transient System Response. (59–255)....177–86
DePriest, J. R. Modern Systems of Traffic Control as Applied to the Seaboard Air Line Railroad Company. (59–251)....168–72
De Remer, J. Grant. Disc....95
Diesel, J. W.; J. Zaborszky. Probabilistic Error as a Measure of Control-System Performance. (59–199)....163–8
Dixon, A. A Recommended Program for Resistance-Welding Instrumentation. (58–210)....199–209
Dixon, C. R.; F. G. Nelson. The Effect of Elevated Temperature on Flash-Welded Aluminum-Copper Joints. (59–1162)....491–5; disc. 495
Duane, J. T. Disc....309
Duane, J. T.; L. J. Yeager. Reliability Analysis for Aircraft Generators. (59–888)....426–34

Duke, R. L.; L. R. Hulls. An Automatic Gage Controller for a 56-Inch Reversing Steel Mill. (58-903)....10-15; disc. 16
Duncan, G. I.; W. A. Rectanus. Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications. (58-118)....16-19

E

Eckert, E. R. G.; J. P. Hartnett, T. F. Irvine, R. Birkebak. The Electrical Analog for Determining Temperature Distribution in Electrical Components. (58-920)....5-10
Eibling, J. A.; D. W. Locklin. Disc....95
Elgerd, Olle I.; W. C. Stephens. Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems. (59-197)....121-7

F

Feiker, G. E.; N. C. Gittinger. Rapid Heating of Dielectric Materials at 915 Mc. (59-30)....35-9
Fiskin, J. M.; W. B. Sisco. Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components. (59-885)....352-6
Forsythe, Stanley D. Rapid Transit is Expanding in Chicago. (56-837)....474-6
Fox, F. C.; S. J. Campbell. Disc....374
Franklin, Gene. Disc....74
Freeman, H. Disc....119
Furumoto, N.; R. M. Buland. Dual-Mode Relay Servos. (59-840)....405-11

G

Gardner, W. F.; E. H. Peters. Disc....506
Gibson, J. E. Disc....53
Gilbert, Edward O. A Method for the Symbolic Representation and Analysis of Linear Periodic Feedback Systems. (59-1183)....512-22; disc. 523
Gittinger, N. C.; G. E. Feiker. Rapid Heating of Dielectric Materials at 915 Mc. (59-30)....35-9
Gleissner, H. Disc....114
Grant, George, III. Load Characteristics of a Submerged-Arc Silicon-Smelting Furnace. (58-1259)....273-5; disc. 275

H

Halacsy, Andrew A. Disc....90
Hamdi-Sepen, D. Transfer Functions of Loaded Synchronous Machine. (58-796)....19-24
Hamilton, W. S. H. Disc....320
Harrington, H. J. Limit-Cycle Stability Study of a Feedback Control System by a New Describing Function Technique. (59-233)....129-33; disc. 134
Harris, L. J.; T. J. Scully, W. N. Stewart. Anode Breaker Testing with a High-Capacity Rectifier System. (58-224)....43-8
Hartnett, J. P.; E. R. G. Eckert, T. F. Irvine, R. Birkebak. The Electrical Analog for Determining Temperature Distribution in Electrical Components. (58-920)....5-10
Herbst, R. J. D. Nichols, F. W. Keay, R. W. Miller. Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer. (59-646)....248-52
Higgins, Thomas J. Disc....394, 444
Hoard, Bert V. Constant-Frequency Variable-Speed Frequency-Make-up Generators. (59-773)....297-301; disc. 303
Hoard, Bert V. Disc....309, 325, 417
Hogwood, E. E., Jr. D-C Electric Drilling Rigs: Application and Operation. (59-1103)....286-9; disc. 291
Hoop, J. G.; D. K. McIlvaine. New Materials for Transformer-Rectifier Unit in High-Temperature Aircraft. (59-883)....450-6
Hopkin, A. M.; R. M. Saunders. Disc....479
Hopkin, A. M.; P. K. C. Wang. A Relay-Type Feedback Control System Design for Random Inputs. (59-218)....228-33
Howard, R. L. New Methods of Simplifying Boolean Functions. (59-91)....134-42; disc. 143
Howell, I. N., Jr.; A. P. Landry. Trends in Ground Bed Design for Cathodic Protection of Underground Structures. (59-820)....456-61; disc. 462
Hulls, L. R.; R. L. Duke. An Automatic Gage Controller for a 56-Inch Reversing Steel Mill. (58-903)....10-15; disc. 16
Hulsey, R. E.; L. L. Kessler. Functional Cycling to Assure Reliability of Aircraft Control Equipment. (59-912)....356-60
Hynes, Lee P. Some Unusual Designs of Electric Resistance Heating. (59-77)....40-3

I

Irvine, T. F.; E. R. G. Eckert, J. P. Hartnett, R. Birkebak. The Electrical Analog for Determining Temperature Distribution in Electrical Components. (58-920)....5-10
Izawa, Keisuke. Disc....394
Izawa, Keisuke; L. E. Weaver. Relay-Type Feedback Control Systems with Dead Time and Sampling. (58-1026)....49-53; disc. 53

J

Jessee, R. D.; W. J. Spaven. Constant Frequency A-C Power Using Variable Speed Generation. (59-810)....411-16; disc. 417
Johannesen, P. R.; F. B. Cox. Application of Switching Transistors and Saturable Reactors in a High-Performance Servo. (58-89)....446-74
Johnson, J. L.; L. E. Moberly. The Effect of Variable High-Altitude Humidity on the Wear of Nondusting Brushes. (59-565)....263-7
Johnson, J. S.; G. I. Moses. Disc....198
Jones, S. J.; R. M. Sills. The Use of Frequency Response Tests in the Analysis of a Foil Mill Automatic Gage Control. (59-78)....59-67
Jury, E. I. Disc....522
Jury, Eliahu I.; G. C. Lendaris. Input-Output Relationships for Multisampled-Loop Systems. (59-682)....375-82; disc. 382

K

Kavanagh, R. J. Disc....120
Kazda, Louis F.; K. Chuang. A Study of Nonlinear Systems With Random Inputs. (59-147)....100-05
Keay, F. W.; R. Herbst, J. O. Nichols, R. W. Miller. Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer. (59-646)....248-52
Keevil, H. W. Detector Car History and New Developments. (59-367)....233-8
Kessler, L. L.; R. E. Hulsey. Functional Cycling to Assure Reliability of Aircraft Control Equipment (59-912)....356-60
Killin, A. M. Disc....275
Kim, W. H.; R. B. Ash, G. M. Kranc. Disc....383
Kinnen, E.; J. Tou. Analysis of Nonlinear Sampled-Data Control Systems:
Part I. (59-828)....386-90; disc. 394
Part II. (59-829)....390-4; disc. 394
Klein, N. Disc....197
Kranc, G. M.; R. B. Ash, W. H. Kim. Disc....383
Ku, Y. H.; A. A. Wolf. A Stability Criterion for Nonlinear Systems. (59-231)....144-7; disc. 147

L

Landry, A. P.; I. N. Howell, Jr. Trends in Ground Bed Design for Cathodic Protection of Underground Structures. (59-820)....456-61; disc. 462
Larson, R. L. A High-Capacity Maintenance-Free Generating System for Motor Coaches. (59-779)....311-16
Lee, Reuben. Disc....89
Lendaris, George G.; E. I. Jury. Input-Output Relationships for Multisampled-Loop Systems. (59-682)....375-82; disc. 382
Lennert, A. E. Practical Considerations of an Ion Propulsion System. (59-833)....326-31; disc. 331
Leonhard, Werner. Elements of Reactor-Controlled Reversible Induction-Motor Drives. (58-1176)....106-14; disc. 114
Lind, E. R.; N. L. Schmitz. Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances. (59-1157)....476-9; disc. 480
Little, D. R.; K. Chen. A Practical Standard Transistorized Optimum Response Controller. (59-838)....337-45; disc. 345
Liu, C. C.; R. Oldenburger. Signal Stabilization of a Control System. (59-219)....96-100; disc. 100
Locklin, D. W.; J. A. Eibling. Disc....95
Lord, H. W.; B. C. Biega. Transient Voltages in Rectifier Transformers. (59-149)....83-9; disc. 90

M

Markowitz, Oscar. Characteristics and Measurement of Ripple in Aircraft Electric Power Systems. (59-879)....418-23; disc. 423
Martin, C. W.; E. W. Bultman. Disc....195
McCausland, Ian. The Optimum Control of Multiactuator Systems. (59-198)....67-73
McIlvaine, D. K.; J. G. Hoop. New Materials for Transformer-Rectifier Unit in High-Temperature Aircraft. (59-883)....450-6
Merriam, C. W., III. Use of a Mathematical Error Criterion in the Design of Adaptive Control Systems. (59-1158)....506-12
Miles, G. F. Disc....461
Miller, R. W.; R. Herbst, J. O. Nichols, F. W. Keay. Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer. (59-646)....248-52
Moberly, L. E.; J. L. Johnson. The Effect of Variable High-Altitude Humidity on the Wear of Nondusting Brushes. (59-565)....263-7
Moses, G. L.; J. S. Johnson. Disc....198
Mulligan, J. H., Jr. Transient Response and the Stabilization of Feedback Amplifiers. (59-1151)....495-503

N

Nakada, Takashi. Disc....100
Nelson, F. G.; C. R. Dixon. The Effect of Elevated Temperature on Flash-Welded Aluminum-Copper Joints. (59-1162)....491-5; disc. 495
Nichols, J. O.; R. Herbst, F. W. Keay, R. W. Miller. Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer (59-646)....248-52
Numakura, Toshio. Disc....134

O

Odok, A. M.; T. M. Soelaiman. Improved D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment. (59-247)....186-93; disc. 198
Oldenburger, R.; C. C. Liu. Signal Stabilization of a Control System. (59-219)....96-100; disc. 100
Owen, T. B. Variable-Speed Constant-Frequency Devices: A Survey of the Methods in Use and Proposed. (59-780)....321-5; disc. 325

P

Pastel, M. P.; G. J. Thaler. Stability Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction. (59-649)....294-7
Penn, W. B.; R. L. Balke, F. M. Precopio. Generator Insulation Systems Development for Hypersonic Aircraft. (59-864)....255-9
Peters, E. H.; W. F. Gardner. Disc....506
Phillips, C. L.; M. H. Yuen. High-Speed Restarting and Protection of Large Synchronous Motors. (59-822)....360-7; disc. 367
Pollard, P. J. A Suppressed-Carrier Signal Generator for Servosystem Instrumentation. (59-1153)....481-5
Precopio, F. M.; W. B. Penn, R. L. Balke. Generator Insulation Systems Development for Hypersonic Aircraft. (59-864)....255-9

R

Randlev, Peter L. Disc....142
Rectanus, W. A.; G. I. Duncan. Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications. (58-118)....16-19
Riaz, M. Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems. (58-917)....25-30
Richards, P. B. Temperature Prediction in Thermal-Lag Equipment. (59-913)....462-5
Rohner, Peter; E. Willihnganz. Battery Impedance: Farads, Milliohms, Microhenrys. (59-823)....259-62
Rosenfein, Allen B.; P. Chandaket. Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading. (59-220)....148-58; disc. 162
Ross, N. V. Three-Phase Induction Heating Coils. (59-256)....291-4

S

Salzer, John M. Disc....382
Saunders, R. M.; A. M. Hopkin. Disc....479
Savo, Theodore A. Disc....158
Schmidt, Stanley F. Application of Continuous System Design Concepts to the Design of Sampled Data Systems. (58-1083)....74-8; disc. 79
Schmitz, N. L.; E. R. Lind. Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances. (59-1157)....476-9; disc. 480
Schuh, N. F.; R. J. Tallent. Solar-Powered Thermoelectric Generator Design Considerations. (59-847)....345-52

Scully, T. J.; L. J. Harris, V. H. Stewart. Anode Breaker Testing with a High-Capacity Rectifier System. (58–224) 43–8
Shaad, G. E.; A. E. Vickery. Transistor Amplistat-Regulated Sectional Drive at Southland Paper Mills. (59–813) ..... 369–74; disc. 374
Shen, D. W. C.; C. F. Chen. A New Chart Relating Open-Loop and Closed-Loop Frequency Responses of Linear Control Systems. (59–647) 252–5
Sheppard, Lyle R. Disc. 461
Shilling, W. J.; R. E. Smith. Disc. 404
Sibley, H. C. Communication Systems for Railway Traffic Control. (59–249) 30–5
Siler, H. N.; J. J. Sonnier. Electrical Features of the Four Corners Pipe Line. (58–1057) 209–23
Sills, R. M. Disc. 15
Sills, R. M.; S. J. Jones. The Use of Frequency Response Tests in the Analysis of a Foil Mill Automatic Gage Control. (59–78) 59–67
Simons, J. S. Disc. 194
Sisco, W. B.; J. M. Fiskin. Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components. (59–885) 352–6
Smith, R. E.; W. J. Shilling. Disc. 404
Smith, Richard T. Disc. 403, 444
Soelaiman, T. M.; A. M. Odok. Improved D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment. (59–247) ..... 186–93; disc. 198
Sollecito, W. E.; D. A. Swann. Computer Evaluation of High-Temperature Aircraft A-C Electrical System Designs. (59–907) ..... 434–44; disc 444
Sonnier, J. J.; H. N. Siler. Electrical Features of the Four Corners Pipe Line. (58–1057) 209–23
Spaven, W. J.; R. D. Jessee. Constant Frequency A-C Power Using Variable Speed Generation. (59–810) ..... 411–16; disc. 417
Staffin, Robert. Executive-Controlled Adaptive Systems. (59–1181) 523–30
Stephens, William C.; O. I. Elgerd. Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems. (59–197) 121–7
Stewart, V. N.; L. J. Harris, T. J. Scully. Anode Breaker Testing with a High-Capacity Rectifier System. (58–224) 43–8
Stineman, R. W. Optimum Reflector-Absorber Geometry for a Solar Generator. (59–869) 332–7
Stineman, R. W. Disc. 331
Stratton, L. J.; K. M. Chirgwin. Variable-Speed Constant-Frequency Generator System for Aircraft. (59–872) ..... 304–09; disc. 309
Stratton, L. J.; K. M. Chirgwin. Disc. 301, 417
Stuart, J. W. Disc 143
Swann, D. A.; W. E. Sollecito. Computer Evaluation of High-Temperature Aircraft A-C Electrical System Designs. (59–907) ..... 434–44; disc. 444

T

Tallent, R. J.; N. F. Schuh. Solar-Powered Thermoelectric Generator Design Considerations. (59–847) 345–52
Tauber, H. Disc. 275
Thaler, G. J.; M. P. Pastel. Stability Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction. (59–649) 294–7
Thomas, J. W.; E. J. Woodfin. Electrified Fibrous Air Filters. (59–680) 276–8
Thompson, G. J.; S. H. Behr. Aluminum Bus for Shipboard Application. (59–654) ..... 230–47; disc. 247
Tiley, G. L.; E. Zucker. Disc. 128
Toffolo, D. S. Disc. 303
Tou, J.; E. Kinnen. Analysis of Nonlinear Sampled-Data Control Systems:
Part I. (59–828) ..... 386–90; disc. 394
Part II. (59–829) ..... 390–4; disc. 394
Trouard, S. E.; E. A. Wagner, Jr. Deep Ground Beds for Cathodic Protection. (59–505) 278–85
Tsang, N. F. Disc. 147
Tudbury, Chester A. Electric and Magnetic Conditions Inside an Induction-Heated Workpiece (59–64) 79–82

V

Vickery, A. E.; G. E. Shaad. Transistor Amplistat-Regulated Sectional Drive at Southland Paper Mills. (59–813) ..... 369–74; disc. 374

W

Wagner, E. A., Jr.; S. E. Trouard. Deep Ground Beds for Cathodic Protection. (59–505) 278–85
Wang, P. K. C.; A. M. Hopkin. A Relay-Type Feedback Control System Design for Random Inputs. (59–218) 228–35
Watkins, William S. Proposed Recommended Practices for Medium-Voltage Motor Controllers for Rubber and Plastics Industries. (59–1112) ..... 504–06; disc. 506
Weaver, Lynn E.; Keisuke Izawa. Relay-Type Feedback Control Systems with Dead Time and Sampling. (58–1026) ..... 49–53; disc. 53
Webb, G. W. Disc. 290
Webb, Robert H. Disc. 238
Wichmann, A. Disc. 193
Willihngans, E.; P. Rohner. Battery Impedance: Farads, Milliohms, Microhenrys. (59–823) 259–62
Withers, W. E. Disc. 289
Wolf, A. A.; Y. H. Ku. A Stability Criterion for Nonlinear Systems. (59–231) ..... 144–7; disc. 147
Woodfin, E. J.; J. W. Thomas. Electrified Fibrous Air Filters. (59–680) 276–8
Wylie, Laurence. Multiple-Unit Operation of Diesel and Electric Locomotives on the Milwaukee Road. (59–492) ..... 316–20; disc. 320

Y

Yeager, L. J.; J. T. Duane. Reliability Analysis for Aircraft Generators. (59–888) 426–34
Yuen, M. H.; C. L. Phillips. High-Speed Restarting and Protection of Large Synchronous Motors. (59–822) ..... 360–7; disc. 367

Z

Zaborszky, J.; J. W. Diesel. Probabilistic Error as a Measure of Control-System Performance. (59–199) 163–8
Zucker, E.; G. L. Tiley. Disc. 128

Design Considerations for Large Industrial Semiconductor Rectifiers

GERALD J. BOWAR
ASSOCIATE MEMBER AIEE

THE SUCCESS in the use of germanium rectifier cells in low-voltage industrial rectifiers since their introduction in 1952 has created a need for higher voltage rectifiers with comparable advantages in efficiency and low initial and maintenance cost. To obtain higher voltages, germanium cells must be operated in series until such time that higher voltages per single cell are made possible.

This paper covers the development of a germanium rectifier rated 250 volts at 7,778 amp (amperes) d-c per cubicle or section. It uses four cells in series. Eighteen of these sections are connected as shown on the one-line diagram represented by Fig. 1 to supply 140,000 amp at 250 volts d-c or a total kw capacity of 35,000 to operate a chlorine line.

The curves or data included in this paper are substantiated by the results of a full-load test applied to the first production section of this type. Although the experience gained here is for a specific germanium rectifier, the basic design principles should be applicable to other ratings and configurations as well as to silicon.

Design Standards

If semiconductors are to be accepted in the large industrial power field, certain basic requirements should be followed or met so as to uphold the standards set forth by these industries. Some of the main considerations are given in the following:

Mechanical:

1. The unit should be sturdily constructed and rugged in appearance.
2. Component parts should be arranged so that they are readily accessible and easily repaired.
3. Bus bars should be rigidly supported and capable of withstanding the maximum short-circuit stresses that could be encountered.
4. Adequate spacing should be maintained between live parts and casing or ground, particularly between d-c busses where it is unlikely that a circuit breaker will be installed.
5. Nontracking or arc-resistant materials should be used for insulating conductors which are connected to high-energy-level d-c loads. Flame-retardent materials should be given serious consideration, particularly when the voltages go above 300 volts.

Electrical:

1. The unit should be conservatively rated to account for the numerous variables affecting current balance in applying large numbers of series-parallel combinations.
2. Fuses or protective devices for the rectifier elements should be carefully selected and co-ordinated with the main overload protection to provide maximum utilization of the equipment.
3. A sufficient number of reliable auxiliary protective devices should be used to protect the equipment adequately and to keep unnecessary shutdowns to a minimum.
4. Safety requirements should be commensurate with the standards of the industry.

Discussion

Rectifier Cubicle

The approach taken for this design was to have an arrangement as shown in Fig. 2. This is a plan view of one subsection of the rectifier where *R*1 through *R*6 represent the six rectifier elements (nine elements in parallel and stacked in a vertical plane in this instance) in a 3-phase bridge circuit. The rectifier elements are mounted in plug-in-type trays to facilitate making the electrical connections to their respective a-c and d-c busses. The trays were designed so that, when fully engaged in the rectifier subsection, they form a duct confining the cooling air to the cell portion of the tray. A very compact arrangement was thus achieved with the ultimate in convenience for repair and maintenance.

To obtain the current rating of 7,778 amp, four of these subsections are required which make up a complete section as shown in Fig. 3. The end subsection houses the d-c disconnect switches and auxiliary controls. General Electric type *LVD* busway was used on the a-c side extending the full length of the rectifier portion of the section. Electrical connections are made to the vertical a-c risers in each subsection, thus paralleling the four subsections. The d-c busses are collected on either side near the bottom of the cubicle and extended into the d-c switch subsection and then brought up through the switches and out the top for the customer's connections.

In this case a separate blower unit, combined with a chilled-water heat exchanger and filter, was furnished by the customer to supply cooling air to the rectifier elements. The air was forced through a duct build into the floor where it entered the rectifier at the bottom and discharged out the top into the room. The air in the rectifier was confined to the cell area of the germanium trays by ducts extending from the lower tray level to the upper portion of the cubicle base. The base was thus pressurized and formed an integral part of the main air duct.

Fig. 4 shows a section without side panels or trays with the internal assembly nearly completed.

Fig. 5 shows a completed section which was taken while the rectifier was being set up for test.

Germanium Trays

The germanium tray consists of four germanium cells connected in series, a fuse, four parallel resistors, and an individual cell-failure indicator. The two lamps on the front of the tray provide indication when either the fuse opens or when one or more of the four cells fail. The germanium tray and cell are shown in Figs. 6 and 7, respectively.

The four cells, when mounted in the tray, are attached to the insulating side panels which also serve as the series-connection point between the cells. The parallel resistors and cell-failure indicator elements are all mounted in the fuse compartment which lies between the front cover and the first germanium cell.

Protective Devices

Twenty-four 20-microfarad capacitors were installed in the rectifier to improve the commutating duty. One of these is connected across each leg of the bridge in each subsection near the center tray.

The auxiliary protective devices in this rectifier section include:

1. A pressure switch sensing the pressure in the main air duct supplying cooling air

---

Paper 58-222, recommended by the AIEE Industrial Power Rectifiers Committee and the Semiconductor Metallic Rectifiers Committee of the Communication and Electronics Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 28, 1957; made available for printing July 1, 1958.

Gerald J. Bowar is with the General Electric Company, Philadelphia, Pa.

The author wishes to express his appreciation to A. Schmidt, Jr., N. J. Lipstein, L. W. Burton, F. W. Kelley, G. R. Lesan, and J. P. Wiles for their assistance in the development and analysis of the design described in this paper.

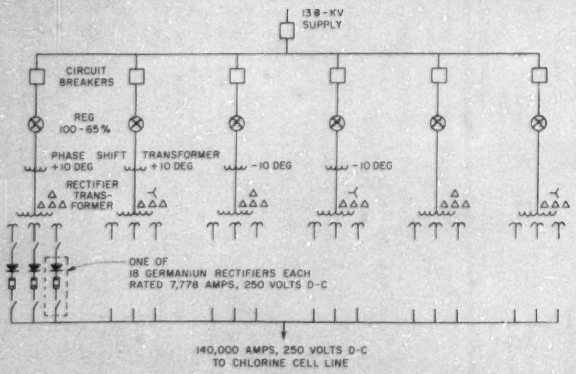

138-KV SUPPLY
CIRCUIT BREAKERS
REG 100-65%
PHASE SHIFT +10 DEG
TRANSFORMER +10 DEG
-10 DEG
-10 DEG
RECTIFIER TRANS-FORMER
ONE OF 18 GERMANIUN RECTIFIERS EACH RATED 7,778 AMPS, 250 VOLTS D-C
140,000 AMPS, 250 VOLTS D-C TO CHLORINE CELL LINE

Fig. 1. One-line diagram of germanium power supply

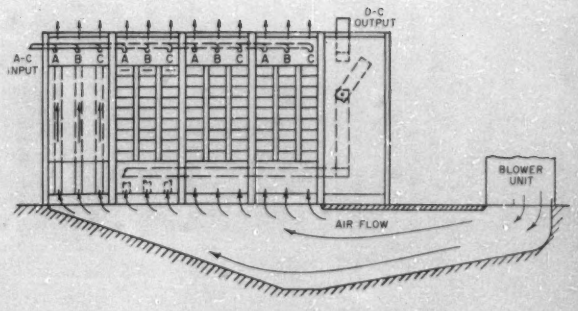


Fig. 3. Elevation view of complete cubicle and air duct in floor

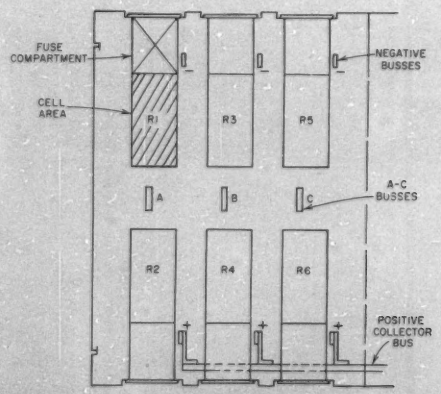


Fig. 2. Plan view of one subsection

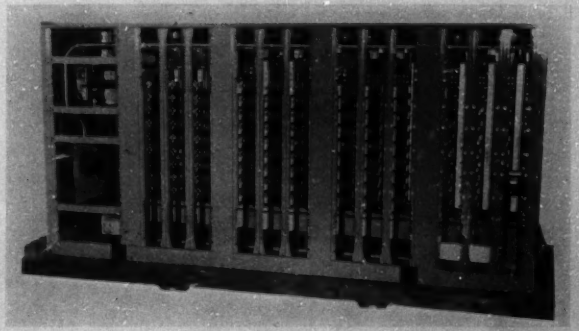


Fig. 4. Rectifier section in final assembly stage

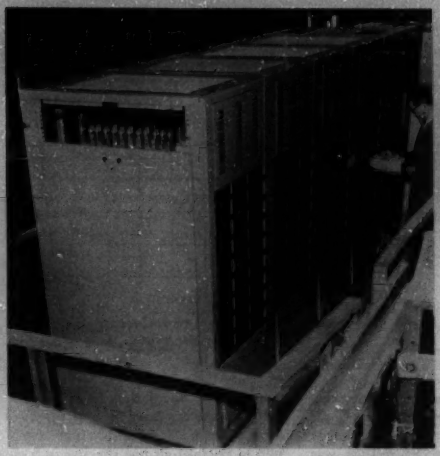


Fig. 5 (left). Rectifier section on load test stand

Fig. 6 (above). Positive germanium tray assembly

to the rectifier elements. This device provides a tripping signal for the main breaker if the pressure, and thus the flow, drop below a set value.

2. An overtemperature device consisting of a resistance temperature detector, located in one of the discharge air paths, which is used in conjunction with an overtemperature relay. This device provides overtemperature protection for the germanium cells when operating at rated load and higher than rated ambient or sustained overloads without compensating for or reducing the ambient.

3. A fuse monitor system provides a tripping signal when the third fuse in any one leg of the bridge opens.

4. An individual cell-failure device which provides indication only when one or more of the germanium cells in a tray fails.

Design Problems

After the general physical arrangement and objectives were established, the design problems centered around the germanium tray and the cubicle structure and bussing.

During the initial design, one of the first problems that had to be resolved was the selection of the number of series cells and a means for insuring equal sharing of the inverse voltage. Four cells were selected to be used in series based on previous experience on low-voltage applications, and 250-ohm resistors were used in parallel with each cell to equalize the inverse voltage. This value was selected by compromising between the maximum voltage the series string could withstand (limiting the voltage to 200 volts maximum per cell) and the maximum watts loss that could be tolerated in the equipment. Under the worst conditions the series string of cells can thus withstand 647 volts and the full-load losses represent approximately 0.3% reduction in efficiency. The maximum surge voltage expected across the string of four cells was 600 to 650 volts which was the anticipated recovery voltage of the fuse. A 150-amp current-limiting-type fuse was used in this case. The tests conducted on the complete section indicated that this peak recovery voltage would be well below 600 volts.

Another basic problem was the selection of the number of cells required in parallel. A number of factors contributed to this selection, namely: volume of air flow, ambient air temperature, air temperature rise in the vertical cooling path, and current unbalance between the parallel elements within the cubicle. The most indeterminate factor in the early design stages was the current unbalance expected when assembling such a large number of cells in a single unit. The final selection was a 6-4-36 cell combination or four cells in series with 36 elements or trays in parallel. This corresponds to approximately a 13% derating in current as compared to the normal rating of these cells when applied to smaller rectifiers used for electrochemical service.

The germanium tray design was an intricate problem within itself primarily consisting of the selection of satisfactory clips for plug-in-type use, providing adequate air seals to be provided between the trays along with matching them to the guide rails provided in the rectifier cubicle. Opposite hand units were made to distinguish readily between the positive and negative side. A forward grading system was established for the complete trays and limits assigned for those to be used in parallel in any one leg. For this production test a single-phase half-wave circuit was used. A 100-amp average short-circuit current was applied with the cooling air controlled to hold the average junction temperatures to 65 degrees centigrade and the voltage drop across the output terminals recorded. A 60-millivolt spread in these readings was allowed for one tray grade. The total number of tray grades was limited to three for each of the positive- and negative-type trays. Each leg of the rectifier bridge, consisting of 36 trays, was thus permitted to use any one of the three grades.

In determining the basic configuration of busses and rectifier elements within the cubicle, an attempt was made to provide a symmetrical arrangement and thus equalize as nearly as possible the impedance drops through the numerous parallel paths. The resistive drop in the vertical a-c and d-c busses was equalized by making them

Fig. 7. Germanium cell assembly

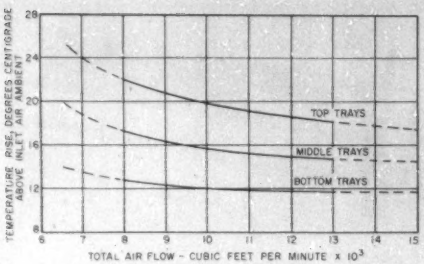

TEMPERATURE RISE, DEGREES CENTIGRADE ABOVE INLET AIR AMBIENT
TOP TRAYS
MIDDLE TRAYS
BOTTOM TRAYS
TOTAL AIR FLOW - CUBIC FEET PER MINUTE x 10³

Fig. 8. Cell temperature rise versus air flow for 8,250-amp load

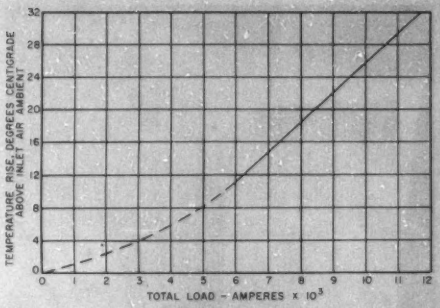


Fig. 9. Cell temperature rise versus load current for constant air flow of 10,500 cubic feet per minute

both of the same cross-sectional area. Although one a-c bus feeds both a positive and negative group of trays, the 120-degree conduction periods are opposite in polarity and are displaced by 60 degrees, so it was assumed that each d-c riser should be of the same cross-sectional area as the a-c riser. A variation in the reactance of the nine parallel paths in one leg connecting these two busses was anticipated, however with the top and bottom trays having a lower reactance and thus carrying more current than the middle trays.

An attempt was also made to balance the impedance drops in the bus bars interconnecting the four rectifier subsections. For the a-c feed, type *LVD* busway was used as previously mentioned. This busway consists of 2-by-$^1/_4$-inch bus bars spaced on 3/4-inch centers and interlaced phase-wise to provide a low-voltage drop. Use of this type of busway reduced the reactive drop present on the a-c side which made it easier to match the d-c busses where there was no reactive drop.

A full-load test conducted on this section revealed the existence of a phase-current unbalance within the cubicle. It became apparent that because of the bus configuration, the commutating reactance between adjacent phases in the four subsections was not the same, with the result that the output current included a circulating component. Inspection of oscillograms indicated that this circulating component originated during commutation and that the currents were steady during the remainder of the conducting period. To correct this condition, small amounts of iron were installed around the vertical a-c busses in varying amounts until a satisfactory balance was achieved.

Full-Load Test

A full-load test was conducted on the first production rectifier section utilizing the facilities in the General Electric Schenectady Works. Here an attempt was made to duplicate the installation at the customer's site. The rectifier was mounted a few feet off the floor above a fabricated air duct and utilized the customer's blower unit as shown in Fig. 5. Electrically, however, the rectifier was connected to a fixed a-c source yielding an open-circuit voltage of nearly 300 volts d-c. The rectifier was loaded through a circuit breaker to some motor-generator sets which pumped power back into the line. Operation under test was thus accomplished by applying full voltage to the section under open-circuit conditions, and then closing the circuit breaker and gradually bringing up the load to the value desired.

Extensive tests were conducted with varying air flows and loads with checks being made on current balance, temperature rise, and general over-all performance. The commutating voltage and current was observed and photographed during which time some experimentation was done with varying capacitances in the circuit. In addition, several fault tests were made by purposely blowing some fuses and taking oscillograms of the currents feeding the fault.

Summary of Test Results

Air Flow

The total air-flow rating for this rectifier section was 10,500 cubic feet per minute at 32 degrees centigrade ambient.

The air-flow distribution through the rectifier trays varied approximately ±3% from the mean in the 24 parallel paths, which was better than expected. The duct itself exhibited a rise in pressure between the blower and rectifier, thereby reducing the speed and power anticipated by the customer and his supplier. Air-leakage tests indicated that about 88% of the inlet air was effective in cooling the germanium cells, which was nearly identical to the predicted value. The air-flow performance thus met or exceeded expectations in all respects.

Current Balance

The total current unbalance expected as a result of these tests is as follows:

1. Current unbalance in parallel trays due to limits of grading ±5.5%.
2. Current unbalance in vertical paths due to cubicle design ±5%.
3. Phase-current unbalance between subsections due to cubicle design ±10%.
4. Subsection current unbalance ±1.5%.

The total unbalance is thus ±22.0%.

Allowing for variation in the test results, the total unbalance should be within ±25% which compares favorably to that total unbalance expected in the early design stages.

Temperature Rise

The average temperature rise above the inlet air ambient on several of the germanium cells as measured by thermocouples is plotted against air flow for the top, middle, and bottom row of trays for a load of 8,250 amp on Fig. 8. The temperature differential between the top and bottom trays closely follows the difference in air temperature as expected. The curves are also quite flat, indicating that variations in air flow should not have an appreciable effect on temperature rise.

Fig. 9 shows a plot of temperature rise versus load which is also of interest. This appears to be a linear function when operating above approximately 3/4 load which is exactly what is expected assum-

ing constant air density which is essentially true for the small differential in temperatures involved.

General Comments

During the 3-week test period, this unit was operating at full load or above for approximately 30 hours. Two cells (864 used in the section) failed during this period, one failure occurring on the initial application of voltage. Both trays with the failed cells were purposely left in the rectifier without any subsequent failures during this test period.

The observations made on the commutating duty indicated that the 20-microfarad capacitors selected were adequate. It was originally thought that individual capacitors across each cell might be needed to improve the commutating duty, but these were found to have no significant effect.

Another test that was tried consisted of removing and reinserting a tray while operating at full load. This was done on several trays with no arcing observed nor any signs of such noted on the contacts upon further examination when the unit was shut down.

One of the final tests was the application of 150% current to one of the four subsections. This load was held for 5 minutes with no ill effects.

Conclusions

The experience gained from this rectifier design shows that considerable emphasis should be placed on attaining a good symmetrical arrangement of components resulting in the best inherent reactance balance in all parallel conducting paths.

These tests have proved that this design of series-parallel combinations of germanium cells is basically sound and it is expected that its performance will meet with good customer acceptance and satisfaction. The tests further prove that design of large blocks of power using semiconductors is only in its infancy. The development of larger cells with greater current-carrying capacity coupled with higher inverse voltages, which are expected with silicon, would further simplify this type of design. These features, combined with the high efficiency of semiconductor elements, make them a natural candidate as the most economical source of d-c power in the future.

The Electrical Analog for Determining Temperature Distribution in Electrical Components

E. R. G. ECKERT
NONMEMBER AIEE

J. P. HARTNETT
NONMEMBER AIEE

T. F. IRVINE
NONMEMBER AIEE

RICHARD BIRKEBAK
NONMEMBER AIEE

SPACE and weight limitations in present-day aircraft necessitate miniaturization of air-borne electric equipment. Such miniaturization generally results in high heat dissipation per unit volume which may lead to local temperatures which exceed the material limitations, causing failure of the component. To avoid such failure, electrical engineers must examine the heat-transfer problem when designing compact electric apparatus. It is the purpose of this paper to present a method of heat-transfer analysis which is relatively simple and straightforward and which utilizes concepts and equipment familiar to electrical engineers. The proposed method is based on the fact that the differential equations describing the flow of heat are identical to those for the flow of electric current; consequently, an electrical analog of the heat-flow problem can be readily constructed and the electrical measurements of current and voltage can be directly translated to the equivalent thermal quantities, heat flow and temperature.

The authors' attention was directed to such cooling problems when they were asked by Jack & Heintz, Inc., Cleveland Ohio, to aid in the determination of the temperature distribution in a high-performance generator coil. The electrical analog was used to solve this particular problem, but it should be emphasized that the method has wide utility.

Nomenclature

e = voltage
E = dimensionless voltage defined by $E = e - e_a / e_w - e_a$
h = heat-transfer coefficient, Btu/hr (British thermal units per hour) square foot, degrees Fahrenheit
i = current flow
i''' = current-density, current per cubic inch
k_t = conductivity of generator coil material
l = electrolytic-tray variable trough length inch
L = maximum dimension of generator coil or analog cross section
n = direction normal to surface
N_q = dimensionless ratio $q'''L^2/k_t(t_w - t_a)$
N_h = dimensionless ratio hL/k_t
q = heat flow, Btu/hr square inch
q''' = internal heat generation, Btu/hr cubic inch
Q = total heat flow, Btu/hr
Q''' = total internal heat generation, Btu/hr
t = temperature, degrees Fahrenheit
T = dimensionless temperature ratio defined by $T = t - t_a / t_w - t_a$
x, y = co-ordinates, inch
X, Y = dimensionless co-ordinates defined as $X = x/L$ and $Y = y/L$

Subscripts

a = air stream
avg = average
e = electrical analog
h = refers to dimensionless parameter in which the heat-transfer coefficient appears
max = maximum
q = refers to dimensionless parameter in which the internal heat generated appears
t = thermal
w = isothermal wall

Analytical Basis for Electrical Analog

The rotor cross section of the Jack & Heintz aircraft generator is shown in Fig. 1 with the generator coil given special designation. An expanded section of the generator coil is given in Fig. 2(A), and the following analysis applies to this

Paper 58-920, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 29-27, 1958. Manuscript submitted March 25, 1958; made available for printing August 21, 1958.

E. R. G. Eckert, J. P. Hartnett, T. F. Irvine, and Richard Birkebak are all with the University of Minnesota, Minneapolis, Minn.

The authors acknowledge the assistance of Roland C. Birkebak in the design of the apparatus and the collection and analysis of the experimental data, the work of F. J. Bradac in the construction of the equipment, and the financial support of Jack & Heintz, Inc.

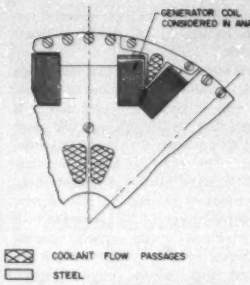
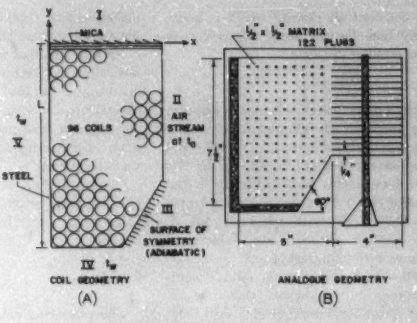
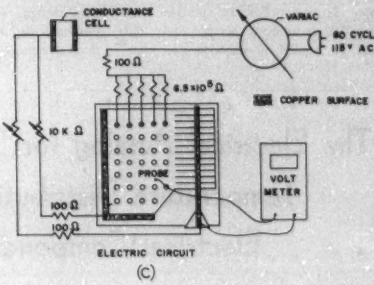


Fig. 1. Rotor cross section

Fig. 2 (right). Sketch of electrolytic analog





geometry. It is assumed that the following data are specified: 1. the heat-generation rate which may be calculated from electrical losses, 2. the temperature of the cooling air, which may be determined by direct measurement, 3. the heat-transfer coefficient for the coolant gas flow over the coil surface which may be estimated from established relations for convection heat transfer if the mass flow rate and the physical properties of the coolant gas are specified, 4. the temperature within the steel material immediately adjacent to the coil (assumed to be constant in the present problem), which may be determined by direct measurement. Starting with this information, the average and maximum temperatures occurring within the coil will now be determined. In addition, the fraction of the total heat generated that flows out through the air-cooled surface will be determined.

It is assumed that the heat flow is steady and 2-dimensional with each coil wire acting as a line source of uniform heat generation. The steady-state 2-dimensional heat-conduction equation with internal heat generation is

$$k_t\left(\frac{\partial^2 t}{\partial x^2}+\frac{\partial^2 t}{\partial y^2}\right)+q'''=0 \qquad (1)$$

Equation 1 can be made dimensionless with

$$X=x/L \quad Y=y/L$$

and

$$T=\frac{t-t_a}{t_w-t_a}$$

where t is the local temperature at any position within the coil, t_a is the air temperature, and t_w is the isothermal wall temperature. The length L is chosen as the maximum coil dimension in the cross section shown in Fig. 2(A). Equation 1 then becomes

$$\frac{\partial^2 T}{\partial X^2}+\frac{\partial^2 T}{\partial Y^2}+\frac{q'''L^2}{k_t(t_w-t_a)}=0 \qquad (2)$$

This equation must be solved subject to the boundary conditions which are imposed on the generator coil, Figs. 1 and 2(A). Surface I is in contact with mica insulation, and it is assumed that no heat flows across this surface; the same assumption is made for surface III because of symmetry of the coil geometry. An air stream at temperature t_a flows past surface II and heat transfer by convection occurs. Surfaces IV and V are in contact with the steel surface of the armature and are at a constant temperature. These boundary conditions can be summarized as

Along face I

$$\frac{\partial t}{\partial y}=0$$

or

$$\frac{\partial T}{\partial Y}=0$$

Along face II

$$-k_t\partial t/\partial x=h(t-t_a)$$

or

$$-\frac{\partial T}{\partial X}=\left(\frac{hL}{k_t}\right)T$$

Along face III

$$\partial t/\partial n=0$$

or

$$\partial T/\partial\left(\frac{n}{L}\right)=0$$

Along face IV

$$t=t_w$$

or

$$T=1$$

Along face V

$$t=t_w$$

or

$$T=1$$

Due to the configuration of the generator coil and the complexity of the boundary conditions, solutions to equation 2 are not readily obtainable. However, the dimensionless heat-transfer parameters involved which control the problem can be obtained directly from equation 2 and the boundary conditions. The dependent dimensionless temperature T is seen to

depend on the independent dimensionless values of the position, heat-generation rate and convective heat-transfer coefficient:

$$f\left(X, Y, \frac{q'''L^2}{k_t(t_w-t_a)}, \frac{hL}{k_t}\right)$$

The solution to this heat-transfer problem can be readily obtained by considering the equivalent electrical problem wherein the same geometry is utilized and the equivalent electrical boundary conditions are imposed. The 2-dimensional steady-state equation describing the electric potential in a conducting material having the shape as shown in Fig. 2(A) may be written:

$$k_e\left(\frac{\partial^2 e}{\partial x^2}+\frac{\partial^2 e}{\partial y^2}\right)+i'''=0 \qquad (3)$$

Equation 3 can be made dimensionlesss when the potential parameter $E=(e-e_a/e_w-e_a)$ and the dimensionless lengths $X=x/L$ and $Y=y/L$ are introduced.

$$\frac{\partial^2 E}{\partial X^2}+\frac{\partial^2 E}{\partial Y^2}+\frac{i'''L^2}{k_e(e_w-e_a)}=0 \qquad (4)$$

The boundary conditions to be imposed on the electrical analog to the coil geometry shown in Fig. 2(A) are

Along face I

$$\frac{\partial e}{\partial y}=0$$

or

$$\partial E/\partial Y=0$$

Along face II

$$-k_e\frac{\partial e}{\partial x}=\left(\frac{k_e}{l}\right)(e-e_a)$$

or

$$-\frac{\partial E}{\partial X}=\left(\frac{k_e}{l}\right)\frac{L}{k_e}E$$

Along face III

$$\frac{\partial e}{\partial n}=0$$

or

$$\partial E/\partial(n/L)=0$$

Along face IV

$$e=e_w$$

or

$$E=1$$

Along face V

$$e=e_w$$

or

$$E=1$$

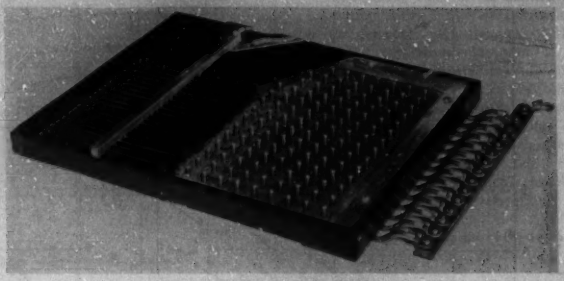


Fig. 3. Actual electrolytic analog

In the electrical analog the thermally insulated surface I and the adiabatic surface III are replaced by electrically insulated boundaries; on surfaces IV and V, a constant-voltage condition is substituted for the isothermal conditions. On surface II the thermal conductance, h, is replaced by the electrical conductance, $k_e/1$, and electric current flows through this conductance to a constant-voltage surface simulating the air stream.

The thermal and electrical equations 2 and 4 are identical, as are their boundary conditions; consequently, a solution to the electrical problem is also a solution to the thermal problem. The following summarizes the analogy.

Independent Parameters:

Thermal

$$\frac{q'''L^2}{k_t(t_w-t_a)}=N_q$$

$$\frac{hL}{k_t}=N_h$$

Electrical

$$\frac{i'''L^2}{k_e(e_w-e_a)}=N_q'$$

$$\frac{(k_e/l)L}{k_e}=\left(\frac{L}{l}\right)=N_h'$$

Dependent Parameters:

Thermal

$$\left(\frac{t-t_a}{t_w-t_a}\right)$$

$$\frac{Q_w}{Q'''}$$

Electrical

$$\left(\frac{e-e_a}{e_w-e_a}\right)$$

$$\frac{I_w}{I'''}$$

The independent electrical parameters, N_q' and N_h', are now varied over the range of interest and the results are directly interpreted in terms of the thermal problem. It may be noted that Q_w/Q''' is also listed with its electrical analog I_w/I''': This quantity is the ratio of the flow of heat (or electric current) through the isothermal faces to the total heat (or current) generated in the coil.

Electrolytic-Tray Analog

The coil geometry was magnified by a factor of 7.5 with the resulting dimensions of the analog given in Fig. 2(B). The electrolytic tray was constructed from a Bakelite material, and the coil geometry milled into the Bakelite plate to a depth of 1/4 inch. Fig. 3 shows an over-all view of the tray. The copper wires of the generator coil which generate heat

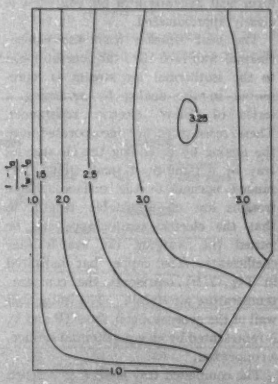


Fig. 4. Isothermal lines for generator coil

$N_q=24$
$N_h=2$

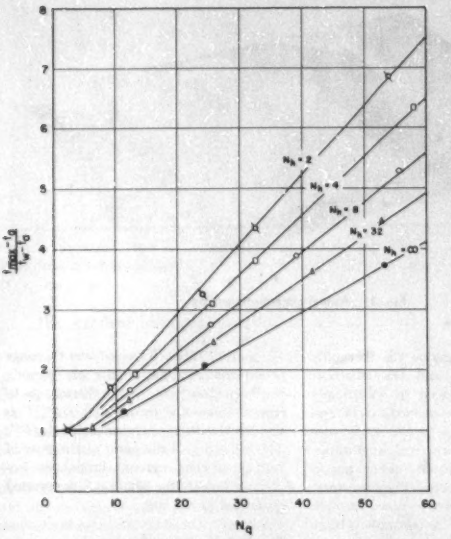


Fig. 5 (above). Maximum core temperature

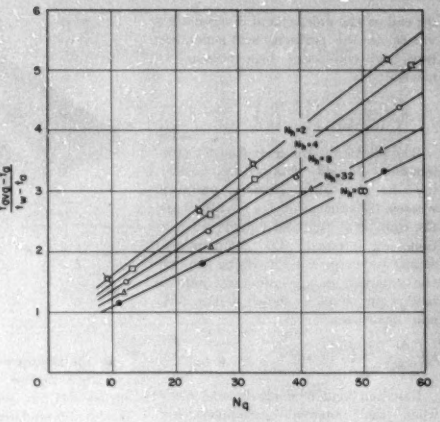


Fig. 6 (above, right). Average core temperature

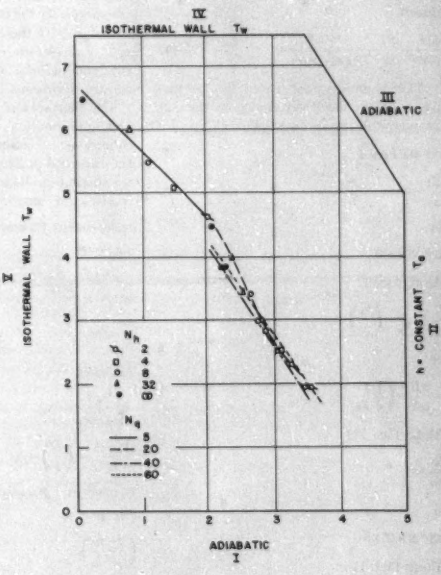


Fig. 7 (right). Position of maximum temperature in coil

uniformly are obtained in the analog by a series of brass plugs shown in Figs. 2 and 3. In order that the current flow through each brass plug be equal, resistors of 6.5 $\times 10^6$ ohms are placed in series with each plug as shown in Fig. 2(C). As a result, the potential drop across the electrolyte is small as compared to the drop across these resistors, and the conditions of uniform heat generation in the coil wires is closely approximated.

The heat transfer from the nonisothermal wall face II of the generator coil to the isothermal air stream is represented in the analog by arranging a series of linear electric resistances. These resistances are incorporated into the analog by extending the electrolytic tray by a number of small troughs arranged normal to the surface. These troughs are of adjustable length, so that the electric conductance can be varied (i.e. varying the heat-transfer coefficient). The copper bar indicated in Fig. 2(B) represents the constant-temperature air stream. The isothermal wall of the generator coil, faces IV and V, is represented by an equipotential surface, a copper bar.

The completed tray is filled to a depth of 1/4 inch with ammonium nitrate, an electrolyte. A conductance cell for measuring the resistivity of the electrolyte was connected directly into the circuit, Fig. 2(C). The dimensions of the cell at the electrolyte surface are 4 by 1 inch and its depth, 1/4 inch. This gives a length to area ratio of 1 inch^{-1} and as a result the resistance measured across the cell gives the specific resistivity ρ_s. (i.e., $\rho_s = 1/k_s$) directly in ohm-inches. The entire electric circuit is shown in Fig. 2(C).

Procedure

The resistivity of the solution was measured at the beginning and at the end of each test to insure that it remained constant. A check was made on the uniformity of current generation at six points in the coil and it was found that

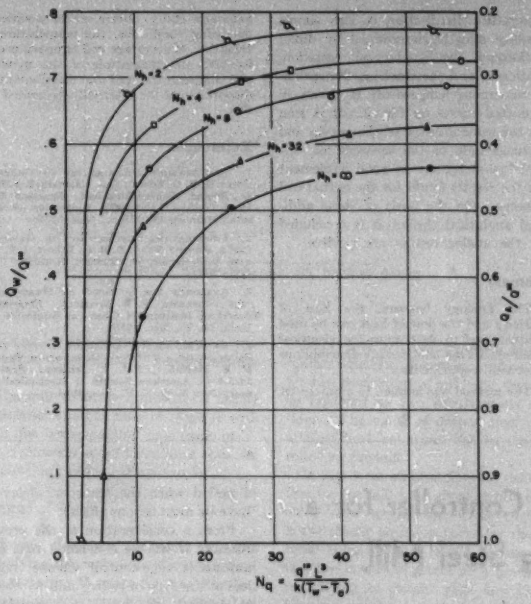


Fig. 8. Heat flow through isothermal wall

this condition was approximated within ±1% if the specific resistivity of the electrolyte was kept less than 60 ohm-inches.

The heat-transfer coefficient represented by the extended tray and the potential difference from the equipotential surface to the equipotential air-stream copper bar were set to desired values. Then the total input current, the current flow to the constant-potential surfaces, and to the air-stream bar were measured. The potential difference from the brass plugs to the air-stream bar were measured with a high-impedance vacuum-tube voltmeter and the equipotential lines throughout the tray were plotted. A typical example is shown in Fig. 4. Sixty-cycle a-c power was used to prevent polarization.

Results

The range of independent parameters, N_q and N_h, investigated for the generator coil were

$N_q = 5$ to 60

$N_h = 2$ to ∞

A prescribed value of N_q was obtained either by varying the current or the potential difference, $e_w - e_a$, while a prescribed value of N_h was obtained by moving the copper bar from the coil boundary. The total current flow, the current flow through the isothermal wall, and the potential distribution were then measured.

The resulting value of the dimensionless maximum core temperature occurring in the generator as a function of the N_q and N_h parameters is shown in Fig. 5. For values of N_q greater than 10, the maximum temperature varies essentially linearly with N_q. A change in the heat-transfer coefficient for small values of N_h is seen to have an appreciable effect on the maximum temperature, while for larger values of N_h, a much smaller change in the maximum temperature is achieved.

Fig 6 shows the variation of the average core temperature defined as

$$\int_0^A t\,dA \int_0^A dA$$

as a function of N_q and N_h. Again, as for the maximum core temperature, for values of N_q greater than 10, the average temperature varies linearly with N_q.

Not only are the maximum and average core temperatures desirable for use in the design of the coil, but also the position of the maximum temperature within the coil is of importance. Fig. 7 gives the variation of the position of maximum temperature for various values of N_q and N_h. For a decreasing heat-transfer coefficient on face II the maximum temperature moves away from face V and when N_h is zero (i.e. face II insulated) the maximum obviously occurs at the junction of faces I and II.

The total energy generated in the coil must be transferred out through the surfaces II, IV, and V. The ratio of the heat flow to the isothermal surfaces IV and V to the total energy generated, Q_w/Q''', is shown in Fig. 8. The fraction of the generated energy which flows to the air-cooled wall is given on the right scale as Q_a/Q'''. At low values of the dimensionless heat-generation rate, N_q, the heat-flow distribution is extremely

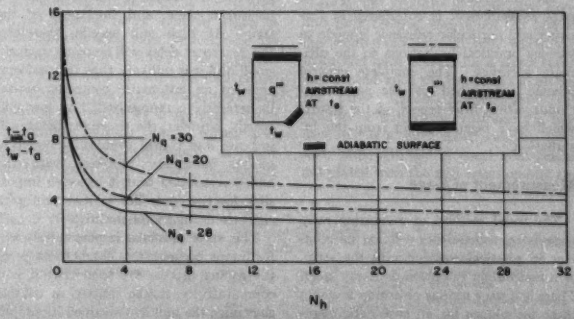


Fig. 9. Heat-transfer results for present coil geometry compared with analytical results for a simpler model

sensitive to slight changes in N_g and N_b. It can be observed that Q_w/Q''' approaches zero for $N_b = 2$ at a finite value of N_g. This indicates that at this condition the air stream just carries away all of the energy generated. Turning to large values of heat generation, say N_g greater than 20, the distribution of the heat flow is found to be essentially independent of N_g and dependent only on N_b.

Finally, an analytical approach was used to verify the results obtained from the analog. A finite difference method was used to calculate the temperature distribution in the coil and the results were in good agreement with the analog values. Additionally, the actual geometry was approximated by a simple rectangular geometry as shown in the upper-right-hand corner of Fig. 9. The temperature distribution in this simple geometry may be determined by direct solution of the conduction equation, equation 2, and the resulting dimensionless maximum temperature is shown in the dotted curves on Fig. 9. It is seen that the general shape of the curves and the magnitude of the answers for this simple geometry are in good agreement with the results found for the actual coil geometry. On the basis of these additional analytical checks, it is concluded that the analog results are reliable.

Summary

1. The analogy between the flow of electricity and the flow of heat can be used to advantage to determine by electrical measurements the temperature distributions in electrical components.

2. The method was applied to a particular generator coil to obtain, under a variety of operating conditions, the temperature distribution, the average coil temperature, the location and magnitude of the maximum coil temperature, and the distribution and disposition of the internally generated heat flux.

References

1. Temperature Distribution in Generator Coils. E. R. G. Eckert, J. P. Hartnett, T. F. Irvine, Jr., Roland, Richard Birkebak. *Technical Report*, Heat Transfer Laboratory, University of Minnesota, Minneapolis, Minn., June 1957.

2. Temperature Distribution in Generator Coils. Report No. 2. E. R. G. Eckert, J. P. Hartnett, T. F. Irvine, Jr., Richard Birkebak. *Ibid.*, Oct. 1957.

3. Analogue for Solution of Heat-Conduction Problems. W. R. Simmons. *Preprint 233*, American Institute of Chemical Engineers, New York, N. Y., Dec. 1955.

4. An Electrolytic Analogue Applied to the Solution of a Thermal Conduction Problem. P. E. McNall, Jr., J. E. Janssen. *Paper no. 54-SA-45*, American Society of Mechanical Engineers, New York, N. Y., June 1954.

An Automatic Gage Controller for a 56-Inch Reversing Steel Mill

R. L. DUKE
ASSOCIATE MEMBER AIEE

L. R. HULLS
MEMBER AIEE

THIS PAPER describes an automatic gage-control system which has recently been put into service on a 56-inch 4-high reversing cold mill. Measurement of gage is achieved by suitable addition of signals proportional to the force tending to separate the work rolls and the unloaded separation between the work rolls. Control of gage is by screwdown and strip tension.

The theoretical basis of this type of gage measurement is considered in this paper and particular reference is made to various practical limitations to the ultimate accuracy. The complete control scheme is described and the paper concludes with a brief report on the results which have been obtained from the installation.

Measurement of Gage

The usual method of measuring strip gage is by using either a flying micrometer in physical contact with the strip, or a noncontact radiation or X-ray gage. There is a very serious objection to using these techniques for an automatic gage-control system, which arises from the physical impossibility of locating the measuring device at the roll bite (i.e. the area where the mill work rolls are in contact with the strip being rolled). If a measurement of gage is not made at the roll bite, there is a transport time or delay time between the time a gage change occurs and the time when the change is detected by the measuring device. The magnitude of the transport delay is dependent entirely upon the distance between the roll bite and the measuring gage, and the speed of the strip. At high mill speeds, therefore, this transport delay will be comparatively small and it is unlikely that the performance of an automatic controller would be seriously compromised. At low mill speeds, however, the transport delay will increase and it will eventually result in serious instability problems with the added restriction that it becomes impossible for the controller to maintain gage right up to the end of the strip.

This state of affairs is particularly unfortunate because the characteristics of the rolling process are such that it is a comparatively simple matter to control gage once the mill has attained its proper rolling speed, and it becomes difficult to do so during acceleration and deceleration, a period when the transport delay will have its most serious effect.

From a consideration of the previous remarks it will be concluded that in an automatic gage-control scheme, regardless of the type of rolling mill to which it is applied, the basic requirement for achieving the best possible performance is a means of measuring gage at the roll bite.

This requirement may be met by using the properties of the rolling mill itself, as a means of measuring output gage. If the mill is regarded as an elastic member obeying Hookes law, and it is assumed that there is no elastic recovery of the material being rolled after it leaves the roll bite, then output gage h_2 is defined by the mill equation as follows:

$$h_2 = F/m + S_0$$

where

h_2 = output gage, inches
F = roll force, defined as the force in tons tending to separate the work rolls
m = elastic coefficient of the mill, tons per inch
S_0 = the separation of the work in inches when F is equal to zero and is controlled by the mill screws

The equation shows that a signal proportional to output gage may readily be

Paper 58-903, recommended by the AIEE Metal Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted April 9, 1958; made available for printing August 27, 1958.

R. L. Duke and L. R. Hulls are with the Canadian Westinghouse Company, Ltd., Hamilton, Ont., Canada.

The authors acknowledge the engineering assistance of R. B. Sims, Davy and United Engineering Company, Ltd., Sheffield, England.

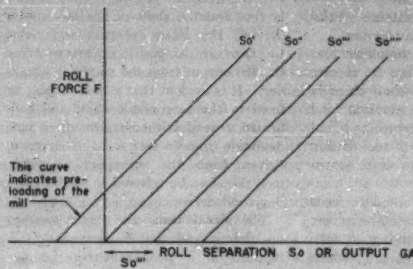


Fig. 1. Mill elastic equation

obtained by the addition of signals proportional to roll force, F, and roll separation, S_0.

The mill equation represents a family of parallel straight lines of slope m with each line corresponding to a value of S_0. Fig. 1 shows a set of these lines obtained by plotting roll force F against h_2.

Control of Gage

Having described the basic principle underlying this technique of measurement it is necessary to relate this to the rolling process, to determine which factors affect the output gage, and to consider the means available for regulating the output gage. For a given setting of the mill screwdowns the output gage will depend upon the following important factors: the ingoing gage, the physical properties of the strip, the front and back tensions, the coefficient of friction between the work rolls and the strip, and the diameter of the work rolls. All these variables may be described on the roll-force h_2 plot by a family of curves known as the plastic curves for the material being rolled. The equations to these curves are very complex and the mathematical derivation involves a somewhat lengthy analysis.[1] However, for the purpose of comprehending the behavior of the mill a simple graphical description of these curves will suffice.

Since both the mill equation and the plastic equation determine output gage for any given value of roll force, they may be regarded as two simultaneous equations. The value of output gage which satisfies both equations will be the gage rolled by the mill.

In Fig. 2, the straight line A is the equation for the mill for a given setting of the screwdown mechanism. The curve B is the plastic equation curve for the material being rolled. It will be seen that the two curves intersect at a point corresponding to output gage h_2. This is simply a graphical method of solving the two simultaneous equations. Now let it be supposed that the properties of the incoming strip change so that there is now a new plastic equation curve C, which intersects the mill line at a point corresponding to the heavier gage h_2'. In other words, a change in the incoming strip has caused the output gage to increase. To correct for this change, the screws may be screwed down so that there is now a new mill line D parallel to the original line but displaced from it so that it now cuts the plastic curve C at a point corresponding to the original output gage h_2. It will be noted that in order to maintain this same gage the roll force has been increased from F to F'. This graphical picture thus provides an illustration of the simple way in which output gage may be controlled by movement of the mill screws. Under normal rolling conditions, therefore, adjustment of screw position provides a very convenient method of controlling gage. From a regulating system point of view it must be remembered that movement of the screws involves the acceleration of reasonably high inertias and hence there is a definite limit to the speed of response on an automatic system using screwdown control only.

When the material is thin the plastic equation curve can become parallel to the roll force axis and at very high values of roll force the curve may even bend toward the region of heavier gage, and consequently the screws will lose their effectiveness as a means of control. Under extreme conditions, not normally encountered in practice, it is possible to reach a point where turning the screws to increase the roll force will actually increase the output gage.

Fig. 3 shows a family of plastic curves for four different values of back tension, together with a mill equation line. It will readily be seen that there are four values of output gage corresponding to each value of tension. Tension, therefore, provides an alternative means of controlling output gage.

The general shape of the plastic curves is such that tension is comparatively ineffective for the control of the heavier gages. Fortunately, tension provides effective control for thin-gage material, a condition under which satisfactory screwdown control may not be possible.

In a single-stand mill where steel is unwound from an entry reel, passed between the mill work rolls, and coiled up on a delivery reel, tension may be very rapidly controlled by regulation of the armature current in the reel-drive machines. Hence tension provides a very fast method of gage control for the thinner gages. Although in a single-stand mill there is no difficulty in regulating front and back tension, this is not the case in the multistand tandem mill where the front and back tension for a given stand will depend upon the conditions in all the the other stands. Consequently, the practice of automatic control in the tandem mill is a much more complex problem, although the physical phenomena associated with the mill equation and the

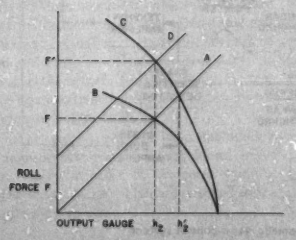


Fig. 2 (left). Determination of output gage from simultaneous solution of mill plastic and elastic equation

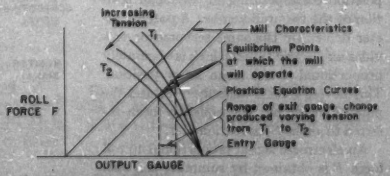


Fig. 3 (right). Effect of changes in tension on output gage

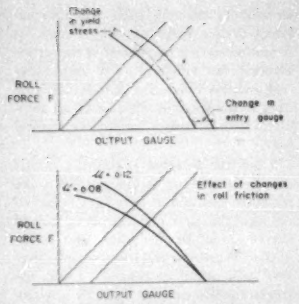


Fig. 4. Effect of changes in roll friction, entry gage, and yield stress on output gage

plastic equation are valid whether the roll bite under consideration be that of a single stand or one of a number of tandem stands.

The principal factors which are going to cause the output gage to change and call for corrective action from the automatic control system are changes in input gage, roll bite friction, and yield stress. Fig. 4 shows a set of plastic curves for changes in these variables. It shows that changes in roll bite friction have considerable effect, and since friction reduces with increasing mill speed it is this fact which introduces the well-known speed effect of output gage reducing during acceleration of the mill.

Description of System

In the 56-inch single-stand reversing cold mill to which the gage-control system has been fitted, position synchros are driven directly from the gear wheels which turn the splined end of the mill screws. This provides a minimum of backlash between the screw and the measuring synchros. A synchro transmitter is driven by one screw and a differential synchro is connected to the other. The receiver synchro is positioned by a small servomotor. The servomotor is driven by an amplifier which is connected in a self-balancing circuit so that the receiver rotor is automatically moved to a position corresponding to the mean position of the two mill screws. A potentiometer driven by the same servomotor provides an electric signal which is accurately proportional to the mean screw position. This device gives a measure of S_0 and the potentiometer has consequently been referred to as the S_0 potentiometer.

An electronic signal proportional to roll force F is obtained by summing the output from the load-cell resistance strain-gage bridges, in a mixing transformer.

The electric summing amplifier and follow-up servomotor, Fig. 5, receive the various input signals, sums them in accordance with the mill equation, and provides a resultant correcting signal which is fed into the payoff reel-tension regulator and the variable-voltage screwdown motor regulator.

The electronic summing amplifier feeds a follow-up servomotor or self-balancing potentiometer circuit, driving a 400-cycle-per-second linear synchro, whose output is fed back into the summing amplifier. Since the servomotor shaft will always turn unless the input signal is zero, the output signal from the linear synchro will always be equal to the sum of the signal inputs to the summing amplifier.

The principal signals which are fed into the summing amplifier are the roll force signal F/m, the roll position S_0, and a set gage-reference signal, which is obtained from a set of decade resistances accurately calibrated in steps of 0.0001 inch. The sum of the signals as given by the output from the linear synchro on the shaft of the follow-up servomotor mill will be equal to the error- or gage-deviation signal (i.e., the difference between the gage reference and the actual gage being rolled). When the reference and rolled gage are in agreement the error is zero and the shaft of the servomotor driving the linear synchro will come to rest at a position corresponding to zero output from the summing amplifier. It is evident that a pointer may be fixed to the servomotor shaft and calibrated to read in thousandths of an inch deviation from the set gage. This servo, driven from the summing amplifier, may therefore be referred to as the gage-deviation servo.

The signal from the linear synchro on the gage-deviation servomotor is rectified in a phase-sensitive detector and amplified in two electronic d-c power amplifiers which feed the screwdown and tension-regulator magnetic amplifiers.

Under automatic control both tension and screwdown are used to control gage. For small gage corrections on thin gages, tension correction assumes the dominant role since the high speed of response enables corrective action to be taken before the slower screwdown regulator has time to operate.

The screwdown control constitutes a position-regulating system and it will, therefore, always tend to reduce the gage deviation to zero, and under these circumstances the tension will always revert to the value set manually by the operator.

To reduce wear by avoiding excessive operation of the screws, particularly during the initial passes, an adjustable deadband is provided in the screwdown loop.

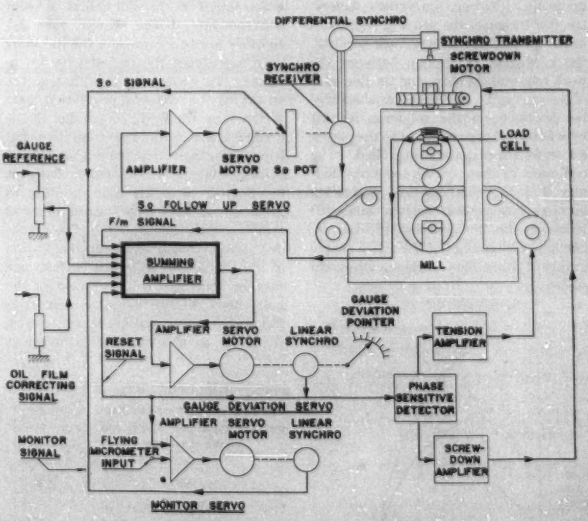


Fig. 5. Block diagram of automatic gage-control system

Any one of four dead-band settings can be selected with a switch calibrated for 0.001, 0.0005, 0.0003 inch, and zero. The gage controller will only operate the screws if the gage deviation exceeds the selected dead-band.

The gain of the screwdown control was adjusted so that a gage deviation of 0.001 inch would drive the screws at their maximum rated speed of 0.515 inch per minute.

Since continuous operation of the tension control presents no wear or maintenance problems there is no deadband in the tension loop.

The gage controller was found to perform satisfactorily using only the unwinding tension for gage-control purposes. When rolling tin-plate gage (i.e., less than 0.015 inch thick) an increase in back tension of approximately 100 pounds per inch width of strip will effect a reduction in output gage of 0.001 inch.

The range of tension control is limited to avoid the possibility of excessive tension, or slack strip. The average tension may be selected by the operator to control strip shape and the gage controller will not cause the pay-off tension to change from this selected value by more than ±5%.

For large corrections the tension amplifier will saturate in its efforts to exceed the permissible 10% band. Under these conditions only the screwdown system is capable of taking the necessary corrective action. On the thicker gages screwdown control is always predominant since under these circumstances tension has little effect upon gage.

It is this feature of dual operation of both tension and screwdown control which is responsible for the successful application of this technique to gage control when rolling tin-plate gages. Attempts to maintain performance by screwdown control alone leads to excessive screw operations resulting from the ineffective efforts of the rather slow screwdown regulator to perform the duties which can so easily be carried out by tension control.

Satisfactory operation of the gage controller does necessitate a certain standard of performance from the screwdown and tension regulators. However, there is no substantial modification necessary to make the system function with the type of equipment normally encountered on a modern rolling mill.

The gage controller also included a monitor system which was found to be necessary in order to overcome some of the inaccuracies which are detailed in the following section.

Factors Affecting the Accuracy of Rolled Gage

In common with all automatic control systems the accuracy limitations imposed upon the controlled variable will depend upon the response of the system and the type of disturbance to which it is subjected. Furthermore, the ultimate accuracy can be no better than the accuracy to which the controlled variable is measured.

In order to achieve satisfactory control of rolled gage, careful consideration must be given to the factors affecting accuracy and compensating devices must be built into the system where necessary.

Temperature Effects

Thermal expansion and contraction of the mill housings, backup rolls, and work rolls will result in a change in the roll opening, S_0, which is not detected by the synchros driven by the screwdown mechanism. The S_0 signal will, therefore, be in error and the gage, as computed from the mill equation, will be incorrect.

This error may be overcome by using a conventional gage-measuring instrument to monitor the automatic gage controller. In the case of the 56-inch mill, the electric deviation signal from a flying micrometer is compared with the signal from the gage-deviation servo and then the difference between the two signals is fed to an integrating servo. The output from the servo then is fed directly to the summing amplifier. The integrator satisfies the necessary requirement for remembering the magnitude of the correcting signal when the monitor gage is not in operation.

Temperature changes in the massive mill housings and rolls take place comparatively slowly and, therefore, there is no necessity for a fast-response monitor. The advantages gained by the elimination of the transport delay are still obtainable when a conventional measuring gage is used to monitor the system slowly.

Change in Oil Film Thickness in the Backup Roll Bearing[8]

Changes in the thickness of the oil film in the backup roll bearings, also result in a change in roll opening which is not detected by the measuring synchros. The thickness of the oil film varies as a function of mill speed but it does not change appreciably with mill load.

When gage measurement is achieved by the technique previously described it is a comparatively simple matter to determine the magnitude of the change in oil film thickness with mill speed.

To do this the mill is run without strip between the rolls, pressure is applied to the rolls, and the speed is varied without changing the position of the screws. At various values of mill speed the decade dials which supply the gage-reference signal are adjusted to reduce the apparent gage deviation to zero. Under these conditions the differences between successive readings on the decade dials give a direct measure of the changes in oil film thickness. Fig. 6 shows the results obtained by running such a test on the mill. It will be noted that the thickness of the oil film increases with speed and the overall result will be a tendency for the gage to reduce with increasing mill speed.

In practice it was found necessary to compensate for the oil film changes by generating a correcting signal, which varies with speed as shown by the dotted line in Fig. 6, and feeding this signal to the summing amplifier.

It is not desirable to rely on the monitor gage to correct for oil film errors as this would necessitate a comparatively fast response monitor which would reintroduce the problem of a transport delay.

Roll Eccentricity

Eccentric rolls will have the effect of introducing a cyclic error into the S_0 signal. The severity of the error due to this cause will be dependent upon the magnitude of the roll eccentricity and the response time of the automatic controller. In practice, normal roll-grinding procedures produce rolls with a sufficiently small eccentricity to enable satisfactory performance to be obtained without an unduly slow controller response.

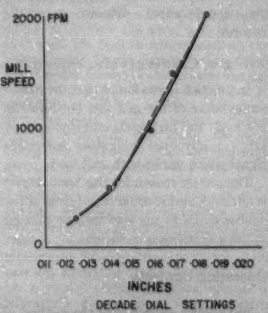


Fig. 6. Result of tests to determine variation of oil film thickness in backup roll bearings with speed for 16½-inch and 53- by 56-inch reversing cold mill

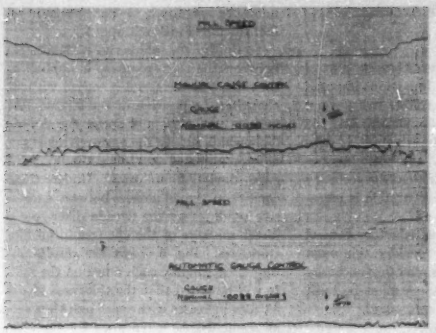


Fig 7. Comparative mill performance on manual and automatic gage control

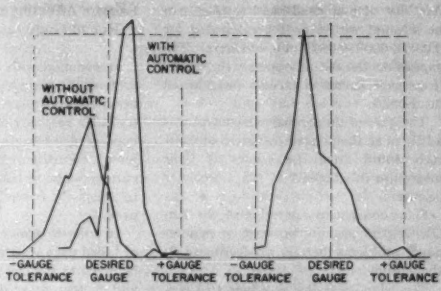


Fig. 8. Gage distribution for tinplate

A—From a single-stand reversing cold mill
B—From a 5-stand tandem cold mill

Speed Effect

It has long been recognized that a rolling mill has an inherent tendency to change gage with speed, the thickness of the strip being rolled decreasing as speed increases, and conversely. It is generally accepted that this effect is due to changes in conditions in the roll bite and changes in the oil film thickness in the roll bearings.

The changes in the roll bite are attributed to changes in friction with speed and they will not result in an error in the gage measurement. However, since these changes will require corrective action to be taken by the automatic controller they constitute a system disturbance which will be introduced every time there is a speed change. Hence even with automatic control there will be a tendency for gage to change during acceleration and deceleration, although the effect will be considerably reduced in comparison with the speed effect without automatic control.

Apparent Changes in Mill Stiffness

In practice it was found that the stress–strain curve of the mill was not linear at low loads and high loads and it was necessary to introduce a further correcting signal which varies with mill load.

The precise reason for this nonlinearity is not clear and it could result from either nonlinearities in the load cells or changes in the mill stiffness coefficient.

Operating Controls

The roller's controls and the method of operation were co-ordinated with existing mill practice on manual control. This insured ready acceptance by the mill crews as the gage control was an addition to, rather than a modification of, established procedure.

Two control desks are used on the mill, one opposite each reel, with much of the control equipment duplicated on each desk. The head roller alternates between desks with each mill reversal so that he is always operating from the delivery side.

Extensions were added to each desk, with the following instruments and control equipment for automatic gage control:

1. Gage setting dials.
2. Gage automatic; manual pushbuttons with indicating lights.
3. Gage-deviation indicator.
4. Mill-load indicators (front and back screws).
5. Monitor selector, which may be either a flying micrometer or X-ray gage. Only the flying micrometer has been used on the 56-inch mill.
6. Gate selector, one desk only.
7. Load-limit selector, one desk only. This sets the value of the mill load for which the automatic controller will lock out and return the mill to manual control.
8. Zero pushbutton to reset S_0 potentiometer.

Operating Sequence

The mill is loaded from a coil box through a leveller and across a hinged table over the reel. The operator may set up the required gage on the delivery desk and on the monitor gage prior to latching the strip into the winding reel. If the gage controller is switched to automatic it will control gage as soon as the mill rolls start turning and the monitor system will be automatically switched on as soon as the strip speed exceeds 150 feet per minute.

During a pass the gage controller and monitor gage may be set up for the next pass on the entry side of the mill and the system automatically switches to the new settings when the mill is reversed.

Before starting the mill it is recom-

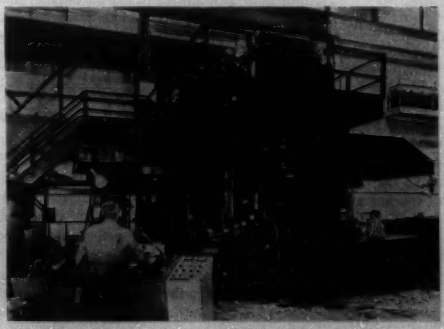


Fig. 9. 16½-inch and 53- by 56-inch, 3,000 - foot - per - minute single-stand reversing cold strip mill with automatic gage control

Fig. 10. Load cell mounted between mill screw and backup roll chock

mended that the operator preset the mill screws to eliminate the rolling of off-gage strip while the automatic controller sets the screws after the mill is running.

The operator is able to resume manual control of the mill at any time.

The only setup procedure required is to reset the S_0 potentiometer after a roll change. To do this the operator brings the mill rolls together at some definite pressure, between 600,000 and 700,000 pounds per screw. Then, by pushing the zero pushbutton, the S_0 potentiometer is unclutched from the synchro and reset by the servomotor.

Performance and Results

The gage control was installed and placed in service in April 1957 operating on screwdowns only. Tension control was incorporated approximately one month later. Fig. 7 shows comparative performance on the mill on manual and automatic gage control.

The distribution curves shown in Fig. 8 indicate that the gage produced on the reversing mill is comparable to that produced on a tandem cold-strip mill. Large-scale sampling tests indicate that 98% of the product is rolled to a gage tolerance better than 0.0004 inch.

Appendix. Mill Installation Data

The gage controller was installed on a 56-inch-wide 3,000-foot-per-minute 4-high single-stand reversing cold mill at Hamilton, Ont., Canada. The mill was designed to roll up to 50-inch-wide strip gages 0.050 inch to 0.006 inch, and up to 36-inch-wide silicon electrical grades.

The mill is driven by a geared twin-motor drive rated 4,340 horsepower, 650 volts, 217/542 rpm, with provision to drive through the 53-inch backup rolls or alternatively through the 16½-inch work rolls. Each reel is driven by a tandem drive rated 2,170 horsepower, 650 volts, 244/732 rpm, with a disconnecting clutch between armatures. The reel-operating tension range is 6,000 to 30,000 pounds with 2-motor drives and 3,000 to 15,000 pounds with single-motor drives.

All mill-regulating systems, reel-tension regulators, mill-generator regulators, mill-motor field regulators, and screwdown regulators utilize 400-cycle magnetic amplifiers. The mill was installed in May 1956 and is shown in Fig. 9. Fig. 10 shows the load cells used for automatic gage control, located between the screw and chock.

The screwdowns are adjustable-voltage drives with a 75-horsepower frame-612 mill motor and a 75-kw 250/500-volt generator on each screw.

References

1. The Rolling of Metals, Vol. 1 (book), L. R. Underwood. John Wiley & Sons, Inc., New York, N. Y., 1953, p. 216.

2. Speed Dependent Variables in Cold Strip Rolling, R. B. Sims. *Journal*, Iron and Steel Institute, London, England, vol. 172, Dec. 1952, pp. 385–95.

Discussion

R. M. Sills (General Electric Company, Schenectady, N. Y.): We agree with the authors' use of a combination of screwdowns and tension for the mill described. In fact, we feel that both reversing and tandem cold-reduction steel mills require a combination of screwdowns and tension to achieve fully automatic control of strip thickness, for similar reasons, although on tandem mills the two systems should be on separate stands. Recent experience in the steel industry indicates wider acceptance of this approach.

We also agree that the gagemeter is a significant contribution to gage control through elimination of transport-time delay. However, it is clear from examination of the thickness-deviation charts that eliminating transport time does not completely eliminate thickness deviations. That is, there are still transient deviations which the control has not eliminated.

From information given in the paper we estimate that the screwdown system has a velocity error coefficient of about 1.7. That is, 0.001-inch error produces 0.0086-inch-per-second screw motion. If there is a 5-to-1 attenuation between screw motion and thickness change, a screw speed of 0.0086 inch per second will produce 0.0017-inch-per-second thickness change, so the velocity error coefficient would be 0.0017/0.001 = 1.7. Thus, crossover frequency will be 1.7 radians per second, and response time (to a small step change) would be close to 2 seconds.

In view of the fact that the screwdown regulators use 400-cycle magnetic amplifiers, we would expect that a regulator with better response could be obtained. We feel that fast response is desirable during the early passes to reduce inherent thickness variations as soon as possible. Is the response limited by motor heating or by stability considerations? If it is stability, what is the source of the troublesome phase shift? That is, is it gear backlash, gagemeter time constants, or the screwdown controls?

In this connection, we note that the gagemeter requires the operation of two servomotors, which tend either to make it slow or to complicate it with high-performance servomotor systems. We suggest that better performance could be obtained by eliminating these servomotors through a change in design.

No mention was made of the gain of the tension system. Can the authors give any figures for this? A relatively high gain and fast response through tension would tend to offset slow response of the screwdowns, particularly on the later passes. The screwdown system would then serve primarily as an automatic positioning control for the screwdowns, to keep tension within range. The authors have implied that this is true.

The paper points out that roll eccentricity will introduce a cyclic error into the S_0 signal. It does not point out that this error will be opposite in direction to the true thickness deviation, and that inherently the control will act to accentuate the thickness error rather than reduce it. Also, we are not sure that all operating people would agree that their normal roll-grinding practices would be satisfactory.

Since it is possible to measure total eccentricity by running the mill and gagemeter with no strip, we wonder whether the authors will indicate the magnitude of eccentricity which results from "normal roll-grinding practice."

The paper mentions four dead bands between 0.001 inch and zero, and indicates that these are used to reduce wear on the screwdowns. Do the authors consider that the excessive motion of the screwdowns was due to variations in entering thickness or to the effects of roll eccentricity? Were there any other factors, such as backlash, motor heating, or stability, that would necessitate the use of a dead band? Can any figures be given as to actual dead band used under various conditions?

In this connection we would like to point out that, although adjustable-voltage screwdowns were used, we see no reason why constant-potential screwdowns would not work almost as well if a dead band is used in either case and control by tension is considered to be a vernier. The screwdown control is basically a position servomotor in which screw position is a function of the integral of gage error. Thus, if a sampling control were used, with ON time proportional to error, nearly equivalent results could be obtained with constant-potential screwdown control.

R. L. Duke and L. R. Hulls: With reference to Mr. Sills' comments, the velocity constant available from the screwdown system would be approximately 2 and is limited by the generator field time constant which is the largest delay in the system. The actual gains existing in practice are very difficult to define because the transfer function of the mill is dependent upon the material being rolled and the particular characteristics of the pass. In this connection, the paper listed in reference 1 attempts to analyze the nature of the mill transfer function. In general, the system gains are set up to provide the best over-all operation for the rolling schedule, although there is no doubt that improved performance could be obtained if the system gains were adjusted both in accordance with rolling speed and the characteristics of the particular pass.

As Mr. Sills suggests, the fast response of the tension control does much to offset the sluggish screwdown operation. Here again the actual gain of the tension regulator cannot be readily specified since it involves the variable gain transfer function of the mill itself.

At low rolling speeds the speed of response of the regulator systems is limited by the eccentricity ripple (average value of about 0.004 inch peak to peak) and it is desirable that the regulator system should not be fast enough to follow the ripple. In the case of a theoretical infinitely fast response system the gage control would put into the strip an eccentricity pattern whose magnitude would be exactly equal to the magnitude of the roll eccentricity. Without any gage control the mill would tend to attenuate the ripple thus producing less ripple in the strip than would be the case if the afore-mentioned theoretical system was used.

At present an experimental eccentricity eliminator is being built and it is hoped that this device will enable the regulator system to be speeded up and result in considerable improvement in performance, particularly for the lower rolling speeds with eccentric rolls.

As far as dead-band settings are concerned, the band is narrowed as the strip becomes thinner. Operation of the screwdowns tends to become excessive on the heavier gages where the tension control is not very effective and in this case the presence of a dead band prevents excessive screw operation with the attendant heating of the screwdown motors. The actual setting of the dead band is under the control of the operators and the particular setting will be chosen which best will suit the rolling conditions. We would agree that there is no reason why good results could not be expected from an ON-OFF type of control operating with constant-potential screwdown control.

It is difficult to eliminate the servomotors from the gagemeter equipment as the system involves accuracies of the order of 1 part in 10,000 which are difficult to obtain without the aid of self-balancing potentiometer circuits and their associated servomotors. An attempt has been made to design an equipment using precision summing amplifiers in place of the servomotors. However, this system was abandoned in favor of the greater accuracy available from servomotors.

REFERENCE

1. A ROLLING MILL ANALOG FOR THE STUDY OF NONLINEAR PROBLEMS ASSOCIATED WITH GAGE CONTROL, L. R. Hulls, K. R. Canfor. *AIEE CP 57-635* (available on request).

Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications

G. I. DUNCAN
ASSOCIATE MEMBER AIEE

W. A. RECTANUS
AFFILIATE MEMBER AIEE

FUTURE ELECTRONIC equipment mounted aboard high-speed military aircraft and guided missiles will have to withstand high altitudes, ultrahigh temperatures, and radiation. Present goals as established by the U. S. Air Force for electronic power transformers are 100,000 feet altitude, 500 C (degrees centigrade) ambient temperature, and intense nuclear radiation. With terminal spacings normally designed for air dielectric strengths at altitudes up to approximately 50,000 feet and temperatures up to 125 C, additional information is needed on breakdown voltages under these new conditions.

The work described in this paper was part of Air Force Contract no. *AF33(616)-3623* for developing temperature- and radiation-tolerant electronic power transformers. It began with a literature search to obtain data on terminal spacing required at high altitudes. While several sources of information were found,[1-4] the data reported were not readily adaptable to the particular design problems. In general, these data were concerned with electrode configurations, electric field conditions, gap lengths, and pressure ranges outside the area of interest.

The objective of this work then became the experimental determination of the voltage-spacing characteristics for the terminals of electronic transformers operating in altitudes up to 100,000 feet and in temperatures up to 500 C. Effects of radiation were not included in this investigation.

The theories pertaining to gaseous breakdown are fully discussed in the works of Loeb[5] and Cobine.[7] At the instant of breakdown, two different types of discharge were observed over the altitude range covered. Up to approximately 70,000 feet the discharge was by sparking; above that altitude it was a glow-type discharge. Evidence of breakdown was also observed on the instruments located in the test circuit. The techniques used for the measurement of voltage were taken from AIEE Standard no. 4.[8]

Data were obtained on the voltage breakdown of needle and sphere electrodes, as well as ceramic-type hermetic terminals as a function of spacing and air density. The tests on needles and spheres were made for 1/4-inch to 4-inch spacings at sea level and at other pressures corresponding to altitudes up to 150,000 feet.

While the needle gap is not an approved method for making voltage measure-

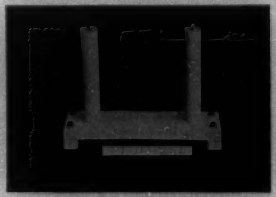


Fig. 1. Needle electrode assembly

Paper 58-118, recommended by the AIEE Air Transportation Committee and the Electrical Insulation Committee of the Communication and Electronics Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing September 12, 1958.

G. I. DUNCAN and W. A. RECTANUS are with the General Electric Company, Fort Wayne, Ind.

The authors wish to acknowledge the help and guidance of G. E. Walter, H. B. Harms, and M. Halleck, as well as the assistance of Howard Ulrich and Gene Galloway in accumulating test data.

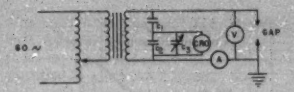


Fig. 2. Schematic wiring diagram

A—Ammeter (Polyranger)
V—Electrostatic voltmeter (Rawson)
CRO—Cathode-ray oscilloscope (Tektronix)
C_1 = 2.2 micromicrofarads
C_2 = 100 micromicrofarads
C_3 = Variable air dielectric capacitor

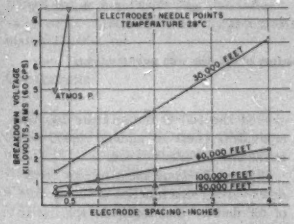


Fig. 3. Dielectric strength of air as a function of needle electrode spacing and altitude

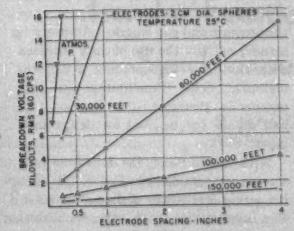


Fig. 4. Dielectric strength of air as a function of a 2-centimeter-diameter sphere electrode spacing and altitude

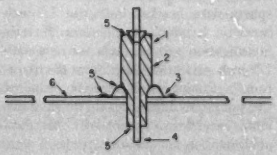


Fig. 5. High-temperature hermetic terminal

1—Alumina ceramic insulation
2—Alumina glaze insulation
3—Nickel-iron alloy gasket
4—Nickel-iron alloy conductor
5—Silver solder seals
6—18-8 type 308 stainless steel

ments in dielectric tests, it is however, a condition encountered when connections are made to transformer terminals. The breakdown voltage on needle electrodes is typical of the erratic results obtained with such connections. The sphere gap method of voltage measurement is a method approved in AIEE Standard no. 4 and represents the best or ideal conditions encountered in spacing terminals with the two sets of data, where the designer has a choice between the worst and best possible conditions. Obviously the data on available terminals narrow the choice.

Needle Point Electrodes

The 60-cycle voltage breakdown of air between needle electrodes at fixed spacings of 1/4 inch, 1/2 inch, 1 inch, 2 inches, and 4 inches was measured in a standard altitude chamber at station level pressure, and at air pressures corresponding to 30,000 feet, 60,000 feet, 100,000 feet, and 150,000 feet altitude. The measurements were made at normal room ambient temperature with the needle electrodes mounted in a ceramic insulator as shown in Fig. 1. Gap lengths were fixed by precision measuring blocks. A new pair of needle electrodes was used for each test condition. Data were determined as the average of three readings at each condition. Needles were 14 gage with a 1/32-inch point radius and 14 degrees total taper. The schematic diagram of the test setup is shown in Fig. 2.

The test results are plotted as a function of altitude and gap length in Fig. 3. They can be summarized as follows: For the range of spacings investigated, the breakdown voltage is a linear function of electrode spacing for any given air density or altitude setting. The slopes of the breakdown versus spacing curves decrease as altitude is increased. At 150,000 feet, the difference in breakdown voltage from a 1-inch to a 4-inch electrode spacing is but 200 volts. For a 1-inch electrode spacing, the breakdown voltage decreases from 2,600 volts at 30,000 feet to 500 volts at 150,000 feet.

Sphere Electrodes

The 60-cycle voltage breakdown of air between standard 2-centimeter-diameter sphere electrodes at 1/4-inch, 1/2-inch, 1-inch, 2-inch, and 4-inch gap lengths was measured in an altitude chamber at station level pressure and at pressures corresponding to 30,000 feet, 60,000 feet, 100,000 feet, and 150,000 feet altitude. The measurements were made at a normal room ambient temperature, and the gap lengths were fixed by precision measuring blocks. Data were an average of three readings for each test condition. The test circuit was the same as shown in Fig. 2.

The test results are given in Fig. 4. They are similar to those experienced with the needle point electrodes except that the slopes of the individual breakdown versus spacing curves for a particular altitude are steeper and are displaced upward. For example, for a 1-inch electrode spacing as altitude increased from 30,000 to 150,000 feet, the breakdown voltage decreased from 16,000 volts to 600 volts. As the electrode spacing is increased from 1 inch to 4 inches at 150,000 feet, the breakdown voltage increases from 600 to 1,300 volts.

Table I. Hermetic-Type Terminals

Terminal	Flashover, Kv*	Outside Diameter, Inches	Terminal to Ground Distance, Inches
A	5.0	0.250	0.375
B	8.0	0.375	0.625
C	16.5	0.750	1.125
D	6.5	0.500	0.375
E	9.0	0.530	0.625
F	17.0	0.530	1.250

*Per MIL-T-27A specification. Values are average based on 60 cycles rms at sea level and 60% relative humidity.

Hermetic Terminals

Tests were made on six different hermetic-type terminals as shown in Table I. The general type of terminal used is shown in Fig. 5. These terminals were selected because of their good electrical properties at elevated temperatures. For testing purposes, ten of each of the terminals were mounted in steel cases. The breakdown voltage of the terminals was measured at station level, and at other air pressure corresponding to 30,000 feet, 60,000 feet, 100,000 feet, and 150,000 feet altitude. The measurements were made at

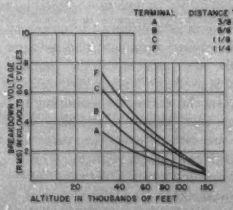


Fig. 6. Breakdown voltage of hermetic terminals as a function of altitude

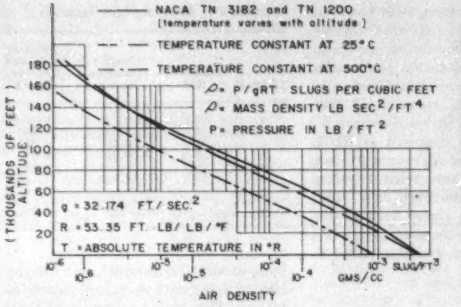

NACA TN 3182 and TN 1200
(temperature varies with altitude)
TEMPERATURE CONSTANT AT 25° C
TEMPERATURE CONSTANT AT 500°C
ρ = P/gRT SLUGS PER CUBIC FEET
ρ = MASS DENSITY LB SEC²/FT⁴
P = PRESSURE IN LB/FT²
g = 32.174 FT/SEC²
R = 53.35 FT. LB/LB/°F
T = ABSOLUTE TEMPERATURE IN °R
ALTITUDE (THOUSANDS OF FEET)
AIR DENSITY
GMS/CC
SLUG/FT³

Fig. 7. Altitude as a function of air density and temperature

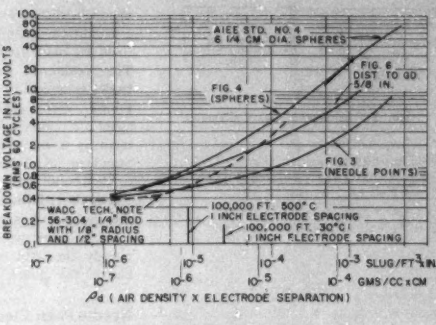


Fig. 8. Comparison of the data with published data

a normal room ambient temperature. The test circuit diagram is shown in Fig. 2.

Data on the breakdown of hermetic terminals are plotted in Fig. 6. Each test point is an average of ten readings. In the presentation of these data, the distance to ground is used as the variable parameter, with the largest spacing being $1^1/_4$ inch. The data are consistent with the sphere and needle point electrode data in that a linear relationship exists between breakdown voltage and spacing for a given altitude setting. A terminal with a $1^1/_4$-inch distance to ground experiences a reduction in breakdown voltage from 14,500 to 1,900 volts as the altitude is increased from station level to 100,000 feet.

Correlation of Data and Significance of Results

The dielectric breakdown strength of air is dependent[7,9] upon its density, the electrode spacing, the electrode material, the electrode configuration, the frequency of the voltage, and the rate of increase of the voltage. For an electric device that has a 100,000-foot altitude and 500 C temperature specification, air density becomes a major variable for the designer to consider in establishing the necessary distance between terminals, and between terminals and ground. If a metallic terminal is used and a single frequency is specified for the device, the remaining two variables that must be seriously considered are electrode configuration and rate of increase of the voltage. This investigation used needle points and sphere configurations that are believed typical of the worst and the best possible situations that can be encountered in transformer applications. A particular type of terminal was used in a third instance to correlate data obtained with a more complex configuration with the simple needle and sphere shapes. The rate of voltage increase recommended by AIEE Standard no. 4 kept this variable a constant throughout the investigation.

The density of the standard atmosphere as recommended by the National Aeronautics and Space Administration (formerly National Advisory Committee for Aeronautics[9,10]) is plotted as the solid line in Fig. 7 up to an altitude of 180,000 feet. It takes into account the variations in temperature of the atmosphere. If a constant temperature of 25 C were maintained, the density-altitude relationship would be that represented by the long dashed line. Then, if the temperature were increased to 500 C, the curve with the short dashes would be obtained. These serve to give a qualitative indication of the effect altitude has on air density and to show that a 500 C temperature can, also, significantly affect air density. Throughout the range of variables considered the 500 C temperature has the equivalent effect of adding 20,000 feet to the altitude specification.

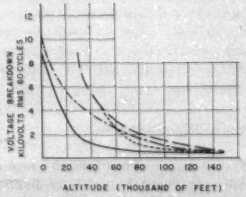


Fig. 9. Comparison of dissimilar electrode configurations of fixed spacings

1/2-inch needles
1/2-inch cylinders
5/8-inch terminals
1/2-inch spheres

It is usually convenient to plot breakdown voltage as a function of the product of air density and electrode spacing to compare and obtain greater utility from gaseous breakdown data. The use of these co-ordinates is based on Paschen's Law[4,7] which states that for a uniform field the voltage just before the discharge of the gap is a function of the product of the pressure and gap length. Accuracy is assured only by the use of data obtained with electrodes that produce a uniform dielectric field. Presentation in this manner is, therefore, a comparison tool and should not be used for design purposes.

With this in mind, the data from four sources are compared in Fig. 8. Besides indicating consistency of data obtained in this investigation with other published data, the range of the density-gap length variable with which we are concerned for transformer applications up to 150,000 feet is a decade or two from the minimum breakdown voltage condition obtainable with a uniform field. The needle and sphere data bracket both the 5/8-inch hermetic terminal data obtained in this investigation and the data for rods with 1/2-inch separation taken from the literature.[11] Cobine establishes the value of the potential difference at which minimum breakdown occurs with his data published[7] on flat-plate electrodes.

Rigorous comparisons and correlations should be made only with data taken for specific spacings and densities. Fig. 9 contains data from this investigation for sphere, needle, and terminal configurations and from the literature for 1/2-inch cylinders. A 1/2-inch gap for simple electrode shapes and a 5/8-inch terminal

were selected for analysis. The cylinder data take an intermediate position between the needle and sphere data. Above 150,000 feet and even somewhat below, electrode shape has little effect on the breakdown voltage. For the 1/2-inch-electrode spacing used, this voltage is approximately 500 volts.

Analysis of the data also shows that spacing is not highly significant above 150,000 feet. At 4 inches, for example, the comparable voltage values range from 700 volts for needle electrodes to 1,200 volts for sphere electrodes. Extrapolation of the data indicates that this latter value would become 700 volts at approximately 170,000 feet.

Conclusions

The data reported in this paper should form a basis for the designer to establish safe voltage breakdown distances between terminals and between terminals and ground for transformer applications that require 100,000 feet altitude and 500 C temperature ratings. The data show that terminal spacings under these conditions must be severely derated from their normal ambient values.

Air dielectric strength data were obtained for the simple electrode shapes such as needle points and spheres and correlated with those taken on a specific terminal type representative of more complex shapes. It was found that: 1. the terminal dielectric system behaves more nearly as sphere electrodes above 70,000 feet and somewhat less ideally below that altitude; 2. electrode separation and electrode shape become less significant as altitudes of 150,000 feet and higher are approached; and 3. the nature of the discharge at breakdown for the electrode shapes evaluated changed from a spark to a glow at approximately 70,000 feet altitude at room temperature.

References

1. Investigation of Porcelain Insulators at High Altitudes, C. V. Fields, C. L. Cadwell. *AIEE Transactions*, vol. 65, Oct. 1946, pp. 656–60.

2. Application of Ideal Gas Theory to the Gaseous Expansion from an Electric Spark, R. B. Edmonson, H. L. Olson, E. L. Gayhart. *Journal of Applied Physics*, New York, N. Y., vol. 25, Aug. 1954.

3. Effect of Altitude on Electric Breakdown and Flashover of Aircraft Insulation, L. J. Berberich, G. L. Moses, A. M. Stiles, C. G. Veinott. *AIEE Transactions*, vol. 63, 1944, pp. 345–54.

4. Space Charge Formation and the Townsend Mechanism of Spark Breakdown in Gases, R. W. Crowe, J. K. Ryaki, V. G. Thomas. *Physical Review*, New York, N. Y., vol. 96, no. 1, Oct. 1, 1954.

5. A Small High-Voltage Bushing Design for High Altitudes, F. J. Vogel, H. A. Hart. *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 761–64.

6. Basic Processes of Gaseous Electronics (book), L. B. Loeb. University of California Press, Berkeley, Calif., 1955.

7. Gaseous Conductors (book), J. D. Cobine-McGraw-Hill Book Company, Inc., New York-N. Y., 1941.

8. Measurement of Voltage in Dielectric Tests. *AIEE Standard no. 4*, Sept. 1953.

9. ICAO Standard Atmosphere. *NACA TN 3182*, National Advisory Committee for Aeronautics, Washington, D. C., 1956.

10. NACA Tentative Upper Atmosphere. *NACA TN 1200*, National Advisory Committee for Aeronautics, 1956.

11. Voltage Breakdown at Low Gas Pressures (High Altitudes), J. W. Ballard, C. B. Goesner. *WADC Technical Note 56-304*, Wright Air Development Center, Dayton, Ohio, July 1956.

Transfer Functions of Loaded Synchronous Machine

D. HAMDI-SEPEN
ASSOCIATE MEMBER AIEE

THE GENERAL STUDY of the transient performance of the synchronous machine is based on Park's[1,2] theory. Various authors have presented useful methods for the study of this subject, and a complete bibliography is contained in reference 3. Problems related to the transient operation of the synchronous machine have been treated as well, aiming at the representation of the synchronous machine for analog studies.[10,11]

Any change in the operating conditions of the synchronous machine causes corresponding changes in electrical quantities. In the case of sudden short-circuiting of an unloaded synchronous machine the current decreases steadily from a maximum to its steady-state value. The variation of the armature current is admittedly related to an apparently variable reactance. Although the value of this reactance is continuously changing,[3] for practical purposes it is taken to be equal to a definite value according to the problem considered. For example, in stability problems the synchronous machine is represented by the transient reactance, just as it is represented by the synchronous reactance in steady-state operation problems.

The transient performance of the synchronous machine can be analyzed directly with the aid of circuit equations, without reference to a variable reactance. Although the theoretical results obtained are readily applicable to other problems, this paper will be concerned with variations of the excitation conditions only, assuming constant torque acting on the shaft of the machine, the speed of rotation corresponding to the nominal frequency. It will be further assumed that the variation of the excitation voltage is relatively small, such as 10% to 20% of its initial value. This limitation is necessary for the approximation admitted in the evaluation of certain terms.

The general case of the transients in the synchronous machine, involving further consideration of torque equations relating to angular position of the rotor, is not within the scope of this paper. In the experimental results reported here, the power-angle variation is deduced from oscillographic records.

In order to arrive at rather simple expressions the amortisseur has not been considered. This affects only slightly the results obtained as response to step variations of the excitation voltage and, in the study of the behavior of the voltage-regulating system given as an example, sufficient approximation is obtained.

The general case of the salient pole machine will be treated. It has been found

Paper 58-796, recommended by the AIEE Feedback Control Systems and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958 and re-presented at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 5, 1957; made available for printing December 15, 1958.

D. Hamdi-Sepen is with the Technical University of Istanbul, Istanbul, Turkey.

The author wishes to acknowledge the use of the facilities of the Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., and the suggestions of Profs. A. E. Fitzgerald and K. L. Wildes; also those of Dr. E. Mishkin, now with the Polytechnic Institute of Brooklyn, Brooklyn, N. Y. Thanks are also due to G. Belfils, Technical Manager of Société ALSTHOM, Paris, France, for the opportunity of working in the research laboratories of this company at Belfort, France, where the experimental results with amplidyne exciter were obtained; and to his staff for its help. The other experimental results were obtained in the laboratories of the Technical University of Istanbul, Istanbul, Turkey.

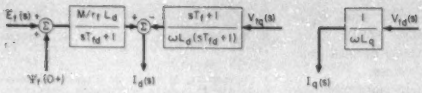
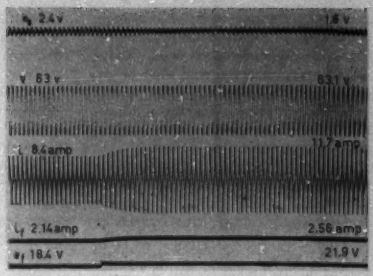


Fig. 1 (above). Block diagram of the synchronous machine



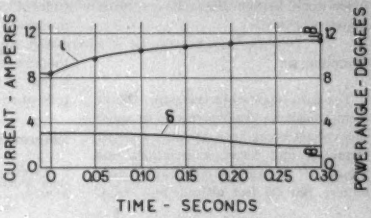


Fig. 2 (right). Results for entirely reactive load with $\phi = 90$ degrees lag

that in round-rotor machines saliency can be of the same order of importance as saturation.[12,13]

Theoretical Considerations

During the operation of the synchronous machine under given terminal conditions the operating conditions of the machine can be changed by the variation of: 1. the voltage applied to the field terminals, and 2. the torque acting on the shaft of the synchronous machine.

In response to the variation of these quantities, two others are changed. They are: 1. the armature current, both in amplitude and phase, and 2. the power angle.

Consider a synchronous machine operating as an alternator connected directly to an infinite bus. If the power of the considered machine is negligible as compared to the power of the system, and if no reactance exists between the terminals of the synchronous machine and the infinite bus, the changing of the excitation of the synchronous machine can change the current delivered but not the voltage at the terminals. Thus the voltage of a loaded machine is not a directly regulated quantity. In case an impedance is connected between the machine terminals and the infinite bus, the voltage at the terminals of the machine will change with the current.

When the synchronous machine is delivering power with current lagging or leading the terminal voltage, any change in the voltage applied to the field circuit terminals will alter both current and power-angle values. A similar changing of current and power angle will occur if the torque acting on the shaft of the synchronous machine is changed, with constant excitation.

Thus in studying the transient performance of the synchronous machine, the voltage applied to the field terminals and the torque acting on the shaft of the machine may be regarded as input functions, while the current and the power angle are output quantities, as considered in feedback control technique.

Impedance Loaded Machine

In a synchronous machine delivering power to an impedance load, with constant torque acting on the shaft, any increase in the voltage applied to the field terminals will increase the armature current. Consequently the RI^2 (resistance impedance) losses in the resistance of the load will increase, with constant driving torque the speed of rotation will decrease, and the system will be operating at a frequency lower than the nominal.

This question will not be considered further in this paper. The case of a pure reactance load or an impedance with negligible resistance is a particular case of the more general problem of the performance of the synchronous machine connected to an infinite bus through an external reactance, in which the voltage of the infinite bus is considered equal to zero.

Synchronous Machine Connected to Infinite Bus

The values of the components of the armature current as given by equations 20 and 23 in Appendix I are related with the condition

$$v_d^2 + v_q^2 = v^2 \tag{1}$$

With constant torque acting on the shaft of the machine, any change in the value of e_f will cause the variation of v_q; v_d and consequently i_q will also be changed, the power angle being

$$\delta = \tan^{-1}\frac{v_d}{v_q} \tag{2}$$

The various terms in equation 21 will be considered. The d-component of the current is dependent upon the following:

1. The voltage $E_f(s)$ applied to the field terminals. If this voltage is varied in the form of a step function to a constant value E_f, this term is $(E_f/s)K_1G_f(s)$ and evaluated as a function of time. If the variation of e_f can not be represented as a simple Laplace transform to be introduced in equation 21 it is possible to replace its variation in a step-by-step calculation by elementary step functions replacing the actual variation of e_f.

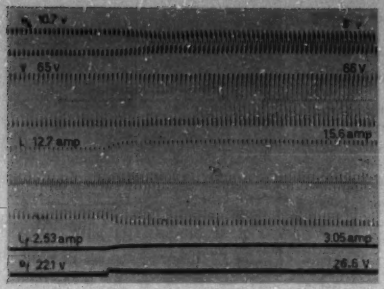
2. The voltage v_q in general is not a known function of time, and is dependent upon δ; the evaluation of the term $V_q(s)K_2G_1(s)$ is possible just in the form of Duhamel's integral in a step-by-step calculation. If the machine is entirely reactive loaded, neglecting the armature resistance v_d would be zero, and v_q constant and equal to v; this term will be $(\overline{V}/s)K_2G_1(s)$ and evaluated. This can still be approximately admitted in operation with lagging power-factor angle about 40 to 30 degrees, for example, according to machine characteristics. The variation of δ and consequently of v_q remains small so that the hypothesis of its constancy is not much in error.

3. The term $\psi_f(0+)K_1G_1(s)$ represents the initial conditions.

As no amortisseur winding in q-axis is considered, neglecting the armature transients, the q-component of the current is directly

$$i_q = v_d/\omega(L_q + L_e) \tag{3}$$

In equations 21 and 23 using $V_q = 0$ and $V_d = 0$ one obtains the expressions of the current components in the case of the reactance-loaded machine considered in the preceding paragraph.



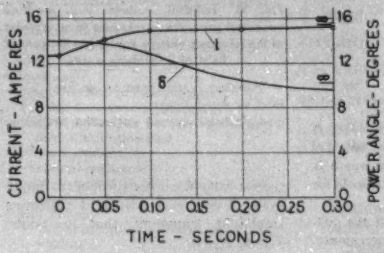
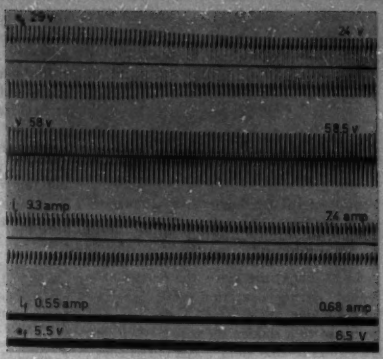
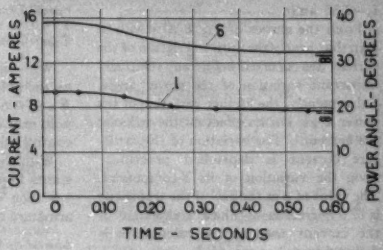


Fig. 3. Results for operation
A (left)—ϕ=47.5 degrees lag
B (right)—ϕ=47 degrees lead





Transfer Functions of Loaded Synchronous Machine

Equations 18(A) and 23 are the basic equations determining the transient operation of the synchronous machine and can be represented in the form of block diagrams as in Fig. 1. $E_f(s)$, $V_{td}(s)$, and $V_{td}(s)$ are input functions, while $\psi_f(0+)$ represents the initial conditions. $I_d(s)$ and $I_q(s)$ are the output quantities.

The block diagram as given in Fig. 1 can be introduced as representing the synchronous machine in any closed-loop system such as the example treated in Appendix II.

Experimental Results

The accompanying oscillograms have been obtained on a 7-kva 110–190-volt 1,500-rpm 50-cycle 3-phase machine. The unsaturated value of X_d is 8.5 ohms and the transient reactance X_d' is 1.7 ohms. We neglect the armature resistance which is 0.279 ohm. M is 0.284 henry; L_f is found equal to 5.59 henrys. r_f is 8.57 ohms, but is larger for currents about 0.5 amp (ampere). Its value is calculated from e_f and i_f as measured. T_f is equal to 0.65 second, and T_{fd}= 0.131 second. The machine was directly connected to the infinite bus.

Tests have been made with terminal voltage about 0.5 per unit, in unsaturated conditions, for quantities related to the d-axis. It is known that the case is different for quantities related to the q-axis.[14]

The variation of the power angle during transients is obtained with the aid of the following arrangement: An auxiliary alternator is mounted on the shaft of the main synchronous machine, with the same number of poles with movable stator. The power angle δ is calculated from phasorial difference voltage $e\delta$ between the voltage of this auxiliary alternator and a voltage taken over a resistor connected between phase and neutral of the infinite bus. These two voltages are regulated to have the same amplitude, and the stator of the auxiliary alternator is rotated so that the phasorial difference between these two voltages is zero when the synchronous machine connected to the infinite bus is running at no load. The stator of the auxiliary alternator remains in this position. During tests these voltages were regulated to 44 volts.

With the foregoing values of the constants, the solution of equations 20 and 23 for the transient current components are

$$i_d = E_f\ 0.87\left(1-\epsilon^{-\frac{t}{T_{fd}}}\right) - V_q\left(0.118+0.472\,\epsilon^{-\frac{t}{T_{fd}}}\right) + \psi_f(0+)\ 6.68\,\epsilon^{-\frac{t}{T_{fd}}}$$

$$i_q = V_d/\omega L_q \qquad (4)$$

In case V_q is variable such as $V_q(t)$, the term containing this voltage component is evaluated in form of Duhamel's integral in a step-by-step calculation.

The value of L_q to be introduced in the last equation is deduced from δ as measured. The saturation effect on this reactance is rather important in this machine.

In Fig. 2 are given the results obtained during operation with entirely reactive load. The initial conditions were v_i=v= 63 volts, i_d=i=8.4 amp, i_f=2.14 amp, e_f=18.4 volts; e_f was suddenly increased to 21.9 volts. The figures on the curves $e\delta$, v, and i in the oscillograms are rms values.

The curves of Fig. 2 give the armature current from oscillogram with calculated points, and the power angle calculated from $e\delta$ as recorded. The errors between calculated and measured values are less than about 3%, partly because of the

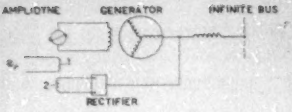


Fig. 4. Voltage-regulating system with amplidyne exciter

reading of the oscillograms which cannot be as precise as the reading of an instrument. The small variation of δ has no effect on the value of the current.

In Fig. 3(A) are given the results obtained during operation with lagging power-factor angle $\phi = 47.5$ degrees, in response to a sudden increase of e_f from 22.1 volts to 26.6 volts. The initial conditions were $v = 65$ volts, $i = 12.7$ amp, $i_f = 2.53$ amp.

From the curves of Fig. 3(A) it will be seen that nearly the entire variation of the current has occurred before any relatively important variation of the power angle. Subsequently, the further variation of the power angle has no effect on the value of the current. The variation of the armature current is dependent practically upon the variation of its d-component. The value of ϕ at the end of the transients is 62.5 degrees lag. The i_q component of the current used in the calculation is deduced from δ as recorded.

The case is entirely different for operation with leading power-factor angle, as seen in Fig. 3(B). The machine was operating with $\phi = 47$ degrees lead, when the voltage at the field terminals was increased from 5.5 volts to 6.5 volts, and the initial conditions were $v = 58$ volts, $i = 9.3$ amp, and $i_f = 0.55$ amp.

The variation of the armature current is entirely dependent upon the variation of its quadrature component and related to the power angle as seen from the curves of Fig. 3(B), though the variation of δ remains related to the variation of ψ_d. Steady-state value of ϕ is 42 degrees lead.

The performance of the voltage regulating system given in Fig. 4 and considered in Appendix II has been tested on a 15-kva 220-volt 39.5-amp 1,500-rpm 50-cycle 3-phase machine. In Fig. 5 the block diagram of the system is given.

For saturated conditions, L_d is 0.0184 henry and M is 0.182 henry. The ratio of field-leakage flux to total field flux is given by manufacturer as 10.5%. Assuming armature leakage of same order, we admit L_{fd}/L_f equal to 0.21; introducing this value in L_{fd} as defined in the Nomenclature with the foregoing values of L_d and M, L_f is found equal to 3.42 henrys. The other constants are

$L_1 = L_2 = 0.91$ henry
$r_1 = 12.7$ ohms
$r_2 = 550$ ohms
$r_Q = 3$ ohms
$T_{12} = 0.073$ second
$T_Q = 0.09$ second
$L_x = 0.00574$ henry
$L_{fdx} = 1.37$ henrys
$r_f = 27.7$ ohms
$T_f = 0.124$ second
$T_{fdx} = 0.0494$ second

With the rectifier used k is 1.35. K as defined in Appendix II is 1,860 and $E_{fo} = 74$ volts. The amplidyne was a 2-pole machine with 125 turns per pole for control fields 1 and 2 (Fig. 4).

With the foregoing values of the constants from equation 32 the incremental variation of the effective value of the armature current is in amp

$$\Delta i = \Delta E_r [4.88 - 1.52 e^{-80t} - 6.9 e^{-7.48t} \sin (11.8t + 29°)] \quad (5)$$

The synchronous machine operating with entirely reactive load, the reference voltage e_r was increased suddenly from 6.9 volts to 7.5 volts.

The initial conditions were $e_f = 93$ volts, $i_f = 3.35$ amp, $i = i_d = 2.4$ amp, $v_t = 210$ volts, and $v = 202.5$ volts.

The oscillogram in Fig. 6 gives the terminal voltage and the armature current, shown in curve a; curve b is calculated with equation 5. The steady-state values after the transients were $e_f = 107$ volts, $i_f = 3.85$ amp, $i = 5.4$ amp, $v_t = 220$ volts, $v = 203$ volts.

The difference between curves a and b is due to the time lag introduced by effects not considered in this analysis, in the variation of the terminal voltage of the synchronous machine actually increasing

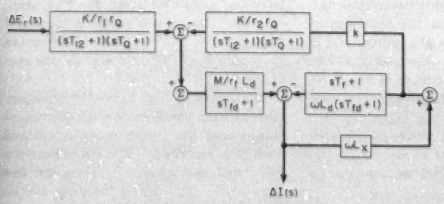


Fig. 5. Block diagram of voltage-regulating system of Fig. 4

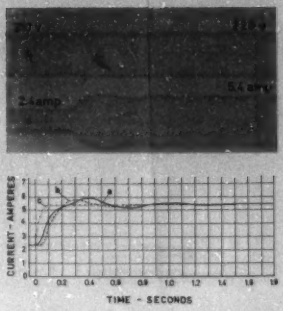


Fig. 6. Terminal voltage and armature current of generator as response to an increase of the reference voltage applied to the control field of amplidyne exciter

a—Armature current from oscillogram
b—Calculated current
c—Calculated current neglecting amplidyne transients

the net control ampere-turns of the amplidyne. Curve c is calculated neglecting amplidyne transients, that is, taking $T_{12} = 0$, $T_Q = 0$.

Conclusions

The expressions of the transient current components have been derived from circuit equations.

The d-component of the transient current is dependent upon the voltage applied to the field terminals and the terminal voltage component corresponding to the resultant flux component along d-axis.

For operation with an entirely reactive load this voltage component is equal to the terminal voltage and is constant for a machine connected to infinite bus. The current in this case depends only upon the excitation voltage.

When the synchronous machine is operating with relatively large lagging power-factor angle so that v_t can be regarded as constant, these conclusions are still valid and the variation of the armature current is dependent practically upon the variation of its component along direct axis.

In case the machine is operating with leading power-factor angle, the variation of both components of the terminal voltage must be taken into account. The armature current variation is almost entirely dependent upon the variation of is quadrature axis component related to the power angle, whose variation however is also dependent upon the variation of the component of the resultant flux along d-

axis. This case is not within the scope of this paper.

An example of use of the transfer functions as derived has been given in the performance calculation of a voltage-regulating closed-loop system with amplidyne exciter, including the consideration of the saturation effects in the amplidyne.

Nomenclature

v = voltage at the infinite bus

v_q, v_d = components of v along q- and d-axis

v_t = terminal voltage of the synchronous machine

v_{tq}, v_{td} = components of v_t along q- and d-axis

i = armature current

i_d, i_q = components of i along d- and q-axis

ϕ = power-factor angle

δ = power angle

ω = speed, electrical radians per second

L_d, L_q = self-inductances per phase

M = mutual inductance between phase and field windings

r = armature resistance

ωL_x = external reactance

e_f = voltage applied to field terminals, also terminal voltage of amplidyne exciter

i_f = field current

L_f, r_f = self-inductance and resistance of field winding

ψ_d, ψ_q = flux linkage components along d- and q-axis

ψ_f = field-winding flux linkage

e_r = reference voltage applied to the terminals of the control field winding of amplidyne

i_1 = current in this winding

L_1, r_1 = self-inductance and resistance of this control field winding

k = ratio of the direct voltage from rectifier to the effective value of the alternating voltage applied

i_3 = current in the field winding of the amplidyne at which the feedback voltage is applied

L_2, r_2 = self-inductance and resistance of this winding

K_E = ratio of the electromotive force induced in quadrature axis circuit of the amplidyne to the effective ampere-turns of control circuits; in case these circuits have the same number of turns, K_E will be in volts per net amp; sum of currents i_1 and i_2

e_Q, i_Q = electromotive force induced and current in short-circuited quadrature axis circuit of amplidyne

L_Q, r_Q = self-inductance and resistance of this circuit

K_{dQ} = ratio of output voltage of the amplidyne to current in short-circuited quadrature axis circuit

K = factor as defined in Appendix II

The various time constants and corresponding inductances are defined as follows:

$$T_f = L_f/r_f$$

$$T_{fd} = \left(L_f - \frac{1.5M^2}{L_d}\right)/r_f$$

$$T_{fdz} = \left(L_f - \frac{1.5M^2}{L_d+L_z}\right)/r_f$$

$$T_{12} = L_1(r_1+r_2)/r_1r_2$$

$$T_Q = L_Q/r_Q$$

$$L_{df}(s) = L_d - \frac{1.5M^2s}{L_fs+r_f}$$

e_3 = voltage equal to the phasorial difference between the voltage of the infinite bus and the voltage of the same amplitude of an auxiliary alternator as defined in this paper

$F(s)$ and $X(s)$ are Laplace transforms of $f(t)$ and $x(t)$, the equations defining the relationship being

$$F(s) \underline{\Delta} \mathfrak{L} f(t)$$

$$X(s) \underline{\Delta} \mathfrak{L} x(t)$$

$f(0+)$, $x(0+)$ are initial values of $f(t)$ and $x(t)$

$\Delta x(t)$ is an incremental variation of $x(t)$, that is $\Delta x(t) = x(t) - x(0+)$

$\Delta X(s)$ is the Laplace transform of $x(t)$

$\bar{X}$ is a constant value of the variable x

Appendix I. Derivation of Transfer Functions of Loaded Synchronous Machine

The equations of Park in their differential form are for the armature winding with $d\theta/dt$ speed of rotation

$$v_{tq} = \frac{d\psi_q}{dt} + \psi_d\frac{d\theta}{dt} - ri_q \tag{6}$$

$$v_{td} = \frac{d\psi_d}{dt} - \psi_q\frac{d\theta}{dt} - ri_d \tag{7}$$

and for the field winding

$$e_f = \frac{d\psi_f}{dt} + r_fi_f \tag{8}$$

In this connection, for a 3-phase machine without amortisseur winding,

$$\psi_d = Mi_f - L_di_d \tag{9}$$

$$\psi_q = -L_qi_q \tag{10}$$

$$\psi_f = L_fi_f - 1.5Mi_d \tag{11}$$

Neglecting the armature transients and resistance, in case of no saturation, and assuming constant speed of rotation corresponding to the nominal frequency, the application of Laplace transformation to the foregoing equations gives

$$V_{tq}(s) = \omega\Psi_d(s) \tag{12}$$

$$V_{td}(s) = -\omega\Psi_q(s) \tag{13}$$

$$E_f(s) = s\Psi_f(s) - \psi_f(0+) + r_fI_f(s) \tag{14}$$

$$\Psi_d(s) = MI_f(s) - L_dI_d(s) \tag{15}$$

$$\Psi_q(s) = -L_qI_q(s) \tag{16}$$

$$\Psi_f(s) = L_fI_f(s) - 1.5MI_d(s) \tag{17}$$

From the foregoing equations one obtains

$$\left(L_d - \frac{1.5M^2s}{L_fs+r_f}\right)I_d(s) = -\frac{M}{L_fs+r_f}[E_f(s)+\psi_f(0+)] - \frac{V_{tq}(s)}{\omega} \tag{18}$$

and

$$I_d(s) = [E_f(s)+\psi_f(0+)]\frac{M}{r_fL_d(sT_{fd}+1)} - \frac{V_{tq}(s)}{\omega}\frac{sT_f+1}{L_d(sT_{fd}+1)} \tag{18A}$$

In case the machine is connected to an infinite bus through an external reactance ωL_x,

$$V_{tq}(s) = V_q(s) + \omega L_xI_d(s) \tag{19}$$

From equations 18 and 19

$$I_d(s) = [E_f(s)+\psi_f(0+)]\times \frac{M}{r_f(L_d+L_x)(sT_{fdz}+1)} - \frac{V_q(s)}{\omega}\frac{sT_f+1}{(L_d+L_x)(sT_{fdz}+1)} \tag{20}$$

This equation can be written

$$I_d(s) = E_f(s)K_1G_1(s) - V_q(s)K_2G_2(s) + \psi_f(0+)K_1G_1(s) \tag{21}$$

with

$$K_1G_1(s) = \frac{M}{r_f(L_d+L_x)(sT_{fdz}+1)} \tag{22}$$

$$K_2G_2(s) = \frac{sT_f+1}{\omega(L_d+L_x)(sT_{fdz}+1)} \tag{22A}$$

If the incremental variations of the variables are considered equation 21 takes the form

$$\Delta I_d(s) = \Delta E_f(s)K_1G_1(s) - \Delta V_q(s)K_2G_2(s) \tag{22B}$$

The quadrature component of the current is

$$I_q(s) = \frac{V_q(s)}{\omega(L_q+L_x)} \tag{23}$$

Saturation Effects

The foregoing results have been obtained for an unsaturated machine. The saturation effects in a synchronous machine can be taken into account, as shown in a previous paper,[14] for quantities related to d-axis by dividing the synchronous reactance by a properly defined factor. If the machine is operating with entirely reactive load no other influence need be considered. Thus L_d will be divided by the saturation factor determined by initial values of voltage and currents. The saturated value of L_d can also be directly deduced from these values. The mutual inductance M would be divided by the saturation factor corresponding to the q-component of the terminal voltage on the no-load characteristic. In a case of an entirely reactive load this component of the voltage is equal to the terminal voltage.

Neglecting the variation of the leakage reactance due to saturation,[15] the saturated value of L_f can be deduced from expressions of leakage reactance such as the transient

reactance. In the case treated as an example in this paper the ratio of field leakage to field flux was known. The saturated value of L_f was deduced from the value of L_M as defined in the Nomenclature.

Appendix II. Calculating the Performance of Voltage-Regulating System with Amplidyne Exciter

The voltage-regulating system considered is given in Fig. 4. The reference voltage e_r is applied to the control field circuit (1) of the amplidyne. The feedback voltage is applied to the second control field (2). This direct voltage is $k_2 v_t$, v_t being the effective value of the terminal voltage connected to the rectifier. The synchronous machine is represented by transfer functions as derived in Appendix I.

Amplidyne performance treatments have been presented by various authors.[12–18] In the following the saturation effects in the amplidyne will be taken into account.

With the incremental variations of the variables, the Laplace transform equation for circuit 1 is

$$\Delta E_r(s) = (L_{11}s + r_1)\Delta I_1(s) - Ms\Delta I_2(s) \qquad (24)$$

and for circuit 2

$$k\Delta V_t(s) = (L_{22}s + r_2)\Delta I_2(s) - Ms\Delta I_1(s) \qquad (25)$$

From these equations, neglecting the leakage flux between circuits 1 and 2, with L_1 equal L_2, one obtains with T_{12} as defined in the Nomenclature

$$\Delta I_1(s) - \Delta I_2(s) = \Delta E_r(s)\frac{1}{r_1(sT_{12}+1)} - k\Delta V_t(s)\frac{1}{r_2(sT_{12}+1)} \qquad (26)$$

Under unsaturated conditions

$$e_Q = K_E(i_1 - i_2) \qquad (27)$$

and for the current

$$\Delta I_Q(s) = \frac{\Delta E_Q(s)}{r_Q(sT_Q+1)} \qquad (28)$$

The amplidyne terminal voltage incremental variation which is also the variation of the voltage applied to the field terminals of the synchronous machine will be, in case of no saturation,

$$\Delta E_f(s) = K_{dQ}\Delta I_Q(s) = \Delta E_r(s)JN(s)/r_1 - k\Delta V_t(s)JN(s)/r_2 \qquad (29)$$

with

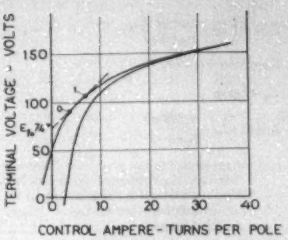


Fig. 7. Load characteristic of the amplidyne

$$JN(s) = \frac{K_E K_{dQ}}{r_Q(sT_{12}+1)(sT_Q+1)} \qquad (30)$$

Replacing this value of $\Delta E_f(s)$ in equation 22B with r_f constant and taking into consideration that

$$\Delta V_{1d} = \omega L_{12}\Delta i_{1d} \qquad (31)$$

the effective value of $\Delta I_d(s)$ which, in a case of an entirely reactive load as considered here, is equal to the incremental variation of the effective value of the armature current will be

$$\Delta I(s) = \Delta E_r(s) \times \frac{K_E K_{dQ} M}{r_1 r_f r_Q \sqrt{2}} \Big/ \left(L_d + L_{12}(sT_{12}+1)(sT_Q+1)(sT_{fd2}+1) + \frac{K_E K_{dQ} k \omega L_{12} M\sqrt{3}}{r_2 r_f r_Q \sqrt{2}}\right) \qquad (32)$$

Saturation Effects

As generally the amplidyne is operating under saturated conditions, the foregoing results need to be modified in order to take into account the saturation effects.

In deriving equation 29 it has been assumed that the amplidyne load characteristic giving the terminal voltage as a function of the resultant ampere-turns of the control circuits was a straight line through the origin.

The actual load characteristic of the amplidyne used in the experimental studies reported in this paper is given in Fig. 7. The point (0) corresponds to the initial conditions, while (1) is the steady-state operating point.

The linear correlation factor between e_f and the resultant control ampere-turns will be taken into account by a new constant factor K instead of $K_E K_{dQ}$, this time along the straight line through points (0) and (1), and a constant value E_{fo} must be introduced.

The terminal voltage of the amplidyne is

$$e_f = E_{fo} + K(i_1 - i_2)/r_Q \qquad (33)$$

defining the factor K which will replace $K_E K_{dQ}$ in equation 32, in the case of saturation of the amplidyne.

References

1. Two-Reaction Theory of Synchronous Machines, Generalized Method of Analysis—Part I, R. H. Park. *AIEE Transactions*, vol. 48, July 1929, pp. 716–30.

2. Two-Reaction Theory of Synchronous Machines—II, R. H. Park. *Ibid.*, vol. 52, June 1933, pp. 352–55.

3. Synchronous Machines (book), C. Concordia. John Wiley & Sons, Inc., New York, N. Y., 1951.

4. Power System Stability (book), S. B. Crary. John Wiley & Sons, Inc., 1947.

5. Voltage Variation of Suddenly Loaded Generators, H. C. Anderson. *General Electric Review*, Schenectady, N. Y., vol. 48, Aug. 1945, pp. 28–33.

6. Electrical Transmission and Distribution Reference Book. Westinghouse Electric Corporation, Pittsburgh, Pa., 1950.

7. The Operational Impedance of the Synchronous Machine, M. L. Waring, S. B. Crary. *General Electric Review*, vol. 35, Nov. 1932.

8. The Effects of Generator Voltage Regulators on Stability and Line Charging Capacity, F. S. Rothe. *Report no. 321*, CIGRE, Paris, France, 1954.

9. Effect of a Modern Amplidyne Voltage Regulator on Underexcited Operation of Large Turbine Generators, W. G. Heffron, R. A. Phillips. *AIEE Transactions*, vol. 71, pt. III, Aug. 1952, pp. 692–97.

10. Fundamental Equations for Analogous Studies of Synchronous Machines, D. B. Breedon, R. W. Ferguson. *Ibid.*, vol. 75, pt. III, June 1956, pp. 297–306.

11. Analogous Computer Representations of Synchronous Generators in Voltage Regulation Studies, M. Riaz. *Ibid.*, Dec., pp. 1178–84.

12. Some Measurements on the Excitation and Load-Angle Characteristics of Large Alternators, J. W. Byrne, D. Carroll. *Report no. 310*, CIGRE, 1956.

13. The Diagnosis of Steam Turbine Alternator Performance by Stroboscopic Method, E. B. Powell. *Report no. 101*, CIGRE, 1956.

14. Saturation Effects in Synchronous Machines, Djabir Hamdi-Sepen. *AIEE Transactions*, vol. 73, pt. III-B, Dec. 1954, pp. 1349–53.

15. A New Approach to the Calculation of Synchronous Machine Reactances—Part II, M. E. Talaat. *Ibid.*, vol. 75, pt. III, June 1956, pp. 317–27.

16. Metadyne Statics (book), J. M. Pestarini. John Wiley & Sons, Inc., 1952.

17. Fundamentals of the Amplidyne Generator, J. L. Bower. *AIEE Transactions*, vol. 64, 1945, pp. 873–80.

18. Transient Analysis of the Metadyne Generator, M. Riaz. *Ibid.*, vol. 72, pt. III, Feb. 1953, pp. 52–62.

Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems

M. RIAZ
MEMBER AIEE

IN A LIST of more than 380 technical problems affecting the defense of the United States prepared by the National Inventors Council (Office of Technical Services, U. S. Department of Commerce) appears the following item under the general heading of Power Supplies: "*741. AC Generators*—Constant frequency, variable speed AC Generators." This is but one indication of the importance attached today to the problem of constant-frequency a-c generation, particularly with regard to air-borne applications. In recent years, many solutions to this problem have been proposed, developed and, in some instances, specific constant-frequency generating systems have been actually utilized in aircraft and guided missiles.

Two basic approaches have been considered: The initial approach had to do with the development of constant-speed drives capable of converting the variable speed of a propeller or turbine prime mover to a constant-speed output shaft which drives an a-c generator. This system is presently employed in many types of aircraft, the control of the constant-speed drives being normally effected on a differential principle by hydraulic means, although pneumatic and mechanical drives have also been suggested. All these systems suffer from a complexity and reliability standpoint, resulting from the close manufacturing tolerances necessary to achieve the desired control at the speeds involved, and from the auxiliary equipment required to supply oil, high-pressure air, and the like.

Despite the relative success of some of these systems, attention has also been given to the possibility of developing all-electric constant-speed drives. The all-electric constant-speed drive assumes the configuration of a variable-speed generator connected to a motor suitably controlled to run at constant speed. One particular arrangement, initially proposed by Pestarini,[1] using d-c metadynes for the generator and motor, offers some interesting possibilities, notwithstanding its large weight and size and the complexities associated with commutators. Another recently suggested arrangement[2] uses a variable-frequency a-c generator to supply the stator excitation of an induction motor, the rotor of which is excited by the difference between the unregulated stator frequency and a standard frequency derived from a relatively low power source. The motor then runs at a constant speed in synchronism with the standard controlled frequency and drives a conventional a-c generator. A basic feature of an all-electric drive is the possible utilization of the variable-speed generated power, whether d-c or a-c, to supply certain types of electrical loads in addition to the constant-speed motor load.

All constant-speed drive systems essentially introduce an extra piece of equipment placed between an input varying speed shaft and the a-c generator. An alternate approach to the problem of constant-frequency generation is to derive this constant-frequency output directly from the variable-speed rotating shaft. Several systems have been suggested along this line of which two kinds can be distinguished; one involves new types of constant-frequency alternators, and the other involves rectification of variable-frequency a-c power followed by its inversion to constant frequency. This inverter-type system is now coming to the fore as a consequence of the recent development of semiconductor devices capable of operating at high power levels. Furthermore, it provides an interesting method of generating constant frequency for those aircraft or missile systems in which the prime source of power is not in the nature of a shaft speed but in the form of direct current.

The majority of the suggested constant-frequency variable-speed a-c generators employs some form of induction machine in conjunction with auxiliary a-c or commutator machines which may, in some cases, form an integral part of the induction machine.[3]

It is with these induction-machine systems arranged to produce constant-frequency generation that this paper is mainly concerned. The object is to determine the basic physical limitations and advantages inherently present in the design of such systems. The approach adopted in this study is based upon consideration of the energy-conversion properties of induction machines.

Electromechanical Energy Conversion in the Induction Machine

Although the theory of the induction motor is now undoubtedly well established and may be found in almost any textbook on rotating electric machines,[4] it is desirable to describe briefly the fundamental characteristics of an induction machine. This review will serve to illustrate the nomenclature and conventions adopted in the analysis. To emphasize the electromechanical energy-conversion aspect, the induction machine as shown in Fig. 1 will be viewed from its three terminal parts through which power is either delivered or received, depending on the particular mode of operation. The relationships between the terminal quantities are given by the performance equations of the induction machine. For steady-state polyphase operation, these equations may be written on a per-phase basis as:

$$V_s = (R_s + jX_s)I_s + ja X_m I_r \tag{1}$$

$$V_r = js\omega X_m I_s + (R_r + js\omega X_r)I_r \tag{2}$$

$$T_e = \frac{\rho a X_m}{\omega_s} \mathcal{R}e\{jI_s^* I_r\} \tag{3}$$

where

V_s, V_r = stator and rotor terminal voltage phasors
I_s, I_r = stator and rotor current phasors
R_s, R_r = stator and rotor resistances
a = equivalent rotor-to-stator turns ratio
s = slip = $(\omega_s - \rho\omega_m)/\omega_s$
ω_s = stator electrical angular frequency
ω_m = rotor angular speed
ρ = number of pairs of poles
X_s, X_r = stator and rotor self-reactances at stator frequency
X_m = component of stator self-reactance due to mutual flux (mutual reactance referred to the stator)
T_e = electromagnetic torque developed, taken to be positive when acting on the rotor in the direction of its rotation

Paper 58-917, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 21, 1958; made available for printing July 2, 1958.

M. RIAZ is with the University of Minnesota, Minneapolis, Minn.

This study was conducted at the Massachusetts Institute of Technology and was supported in part by Wright Air Development Center under Contract *AF 33(616) 3984*.

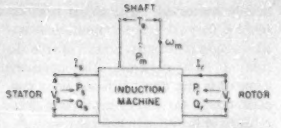


Fig. 1. Block diagram of the induction machine as viewed from its terminals

x_s = stator leakage reactance = $X_s - X_m$
x_r = rotor leakage reactance = $X_r - a^2X_m$

It is often convenient for numerical calculations to express the performance of an induction machine in terms of an equivalent circuit. The particular form of the equivalent circuit illustrated in Fig. 2 may be derived directly form equations 1 and 2 using the concept of an ideal induction machine having a voltage ratio of $as/1$ and a current ratio of $1/a$. This equivalent circuit made up of two subnetworks is convenient because it separately exhibits each of the stator and rotor circuits and so conforms with the terminal representation given in Fig. 1. The polarities and directions of the voltages, currents, and powers are indicated in Fig. 2. Power will be taken to be positive when delivered to the induction machine. In order to stress the power conversion characteristics of the induction machine, the stator and rotor resistances are assumed to form part of the external circuits connected to the machine and attention is focused on the electrical terminals defined by the voltages E_s and E_r in Fig. 2. The per-phase power expressions are defined as

$$\Pi_s = P_s - |I_s|^2R_s = \mathcal{R}e[I_s^*E_s] = |I_s||E_s|\cos\theta_s \quad (4)$$

$$\Pi_r = P_r - |I_r|^2R_r = \mathcal{R}e[I_r^*E_r] = |I_r||E_r|\cos\theta_r \quad (5)$$

$$Q_s = \mathcal{I}m[I_s^*E_s] = |I_s||E_s|\sin\theta_s \quad (6)$$

$$Q_r = \mathcal{I}m[I_r^*E_r] = |I_r||E_r|\sin\theta_r \quad (7)$$

$$P_m = -T_e\omega_m = -T_e(1-s)\frac{\omega_s}{p} \quad (8)$$

where P and Π indicate real or active power and Q denotes reactive power.

Carrying out the operations implicit in the foregoing definitions through the use of equations 1, 2, and 3 yields

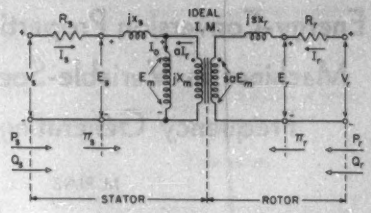
$$\Pi_s = aX_m\mathcal{R}e[jI_s^*I_r] \quad (9)$$

$$\Pi_r = saX_m\mathcal{R}e[jI_s I_r^*] = -s\Pi_s \quad (10)$$

$$P_m = -(1-s)\frac{\omega_s}{p} \times \frac{paX_m}{\omega_s}\,\mathcal{R}e[jI_s^*I_r] = -(1-s)\Pi_s \quad (11)$$

The addition of these three equations gives

Fig. 2. Electrical equivalent circuit of the induction machine



$$\Pi_s + \Pi_r + P_m = 0 \quad (12)$$

which is an expression of the principle of conservation of power.

The relation between stator and rotor reactive powers assumes the form

$$Q_s + \frac{Q_r}{s} = Q_0 \quad (13)$$

where

$Q_0 = |I_s|^2x_s + |I_r|^2x_r + |I_0|^2X_m$
= sum of reactive powers needed to establish leakage and magnetizing flux

This also expresses a conservation principle in terms of reactive powers.

Equations 10, 11, and 13 are of fundamental significance in characterizing the steady-state properties of the induction machine. The electrical properties may therefore be summarized by considering the four basic transformations performed by this electromechanical energy converter:

1. A frequency transformation between stator and rotor given by

$$\omega_r = s\omega_s \quad (14)$$

where s is the slip.

2. A transformation of stator to rotor voltages given by the ratio 1 to as.

3. A transformation of the stator and rotor active powers determined by slip only, as expressed by equation 10.

4. A transformation of reactive powers as given by equation 13.

Characteristics of Induction Machines

The characteristics that govern any particular mode of operation of the induction machine can be derived as special cases of the general analysis presented in the previous section. The examination of what happens to the power flow leads to distinguishing six basic types of operation. These six types are illustrated in Fig. 3; their distinctive features are solely determined by the value of the slip and the direction of energy conversion. A brief discussion of each of these possible configurations follows next.

CASE A: $s<0$; $T_e>0$

This case corresponds to that of an induction motor operating above synchronous speed when, in addition to the

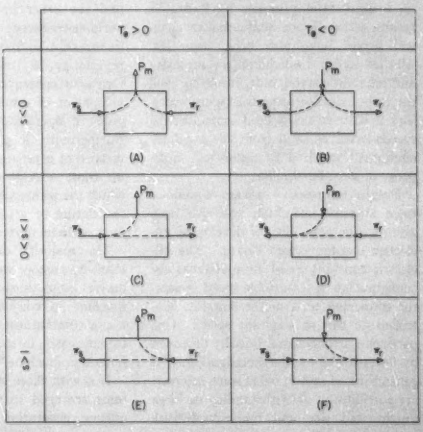


Fig. 3. Power transfer in induction machines

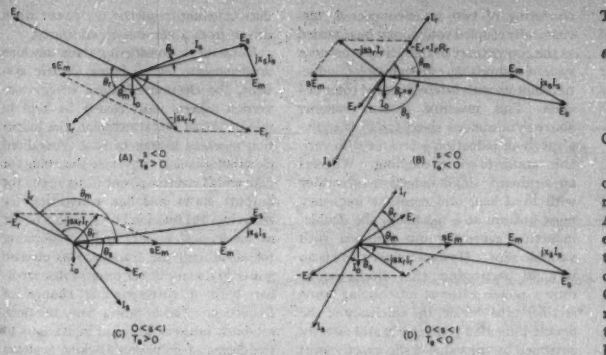


Fig. 4. Phasor diagrams corresponding to the different modes of operation of an induction machine

power supplied to the stator, slip-frequency power is fed to the rotor. For a fixed slip frequency, the motor speed is nearly independent of load. By shifting the phase of the injected rotor voltages, the power factor at which the motor operates can be controlled. This can be seen from the relation.

$$\theta_s = \tan^{-1}\frac{Q_s}{\Pi_s} = \tan^{-1}\frac{Q_0 - (Q_r/s)}{-\Pi_r/s}$$

in which $s<0$. For instance, leading-power-factor operation results when Q_r is negative and greater than $|sQ_0|$. This condition of operation is described by the phasor diagram of Fig. 4(A).

CASE B: $s<0$; $T_e<0$

This is the case of the mechanically driven induction generator which delivers active power to external loads connected to the stator and rotor. Included in this power are the I^2R losses in both stator and rotor. The lagging reactive power demands ($Q>0$) of the loads as well as the reactive power Q_0 of the machine must be supplied by suitable external circuits which usually assume the form of capacitors or overexcited synchronous generators conected in parallel to the induction generator.

The conventional induction generator has a short-circuited rotor ($\Pi_r = I_r^2 R_r$). It is characterized by a load and power factor operation that is completely determined by the value of slip. If operated in parallel with synchronous machines, the induction generator has its voltage and frequency fixed by these machines. If used in an isolated system, the induction generator can provide its own excitation when a suitable capacitor bank is connected across its terminals, thereby establishing a resonant circuit condition. The phasor diagram governing this situation is shown in Fig. 4(B). Only the active component of stator current $I_s \cos\theta_s$ (neglecting stator resistance effects) can be supplied by the generator to a load while the lagging reactive component of stator current has to be furnished by external capacitors or synchronous machines. A factor expressing the efficiency of power conversion in the induction generator may be defined as

$$\eta_g = \frac{\text{stator power output}}{\text{mechanical power input}} = \frac{-\Pi_s}{P_m} = \frac{-\Pi_s}{-(1-s)\Pi_s} = \frac{1}{1-s} \quad (s<0)$$

The limitation of the ordinary induction generator to operate at a fixed slip for a constant frequency and fixed load may be overcome by injecting suitable voltages in the rotor circuit.

CASE C: $0<s<1$; $T_e>0$

This is the case of the induction motor operating below synchronous speed. In general, the rotor slip-frequency power may supply, in addition to the rotor $I_r^2R_r$ losses, some external active or passive circuits which are so adjusted as to control speed and power factor. The phasor diagram for this mode of operation is given in Fig. 4(C).

The ordinary induction motor has a short-circuited rotor ($E_r = -I_r R_r$). Its power conversion efficiency is defined as

$$\eta_m = \frac{\text{mechanical power output}}{\text{stator power input (neglecting stator losses)}} = \frac{-P_m}{\Pi_s} = 1 - s \quad (s>0)$$

The power factor angle is

$$\theta_s = \tan^{-1}\frac{Q_s}{\Pi_s} = \tan^{-1}\frac{sQ_0}{-\Pi_r} = \tan^{-1}\frac{sQ_0}{I_r^2 R_r} \quad (s>0)$$

CASE E: $s>1$; $T_e>0$

This also defines a motor operation completely equivalent to Case C, with the role of stator and rotor interchanged. Assuming the same sequence of polyphase excitation, the motor in Case E is excited through its rotor and turns in the opposite direction to the motor operating as in Case C with stator excitation. Because of this reversal of rotation, the slip being customarily defined with respect to the stator becomes greater than unity. Clearly the slip with respect to the rotor is $1/s$.

CASE D: $0<s<1$; $T_e<0$

Here, input power is delivered mechanically to the shaft and electrically to the rotor while the stator supplies electric power to an external load. The phasor diagram representing these conditions is given in Fig. 4(D). This generator operation is characterized by an energy balance in which s-per-unit power is delivered electrically to the rotor circuit and $1-s$ per unit is supplied mechanically to the shaft to produce one-per-unit stator output power. Power amplification is therefore achieved as determined by the ratio

$$A = \frac{\text{electric power output}}{\text{electric power input}} = \frac{-\Pi_s}{\Pi_r} = \frac{1}{s}$$

which is also the frequency-transformation ratio of the induction machine (equation 14).

CASE F: $s>1$; $T_e<0$

This case also refers to a generator operation similar to Case E but with the roles of stator and rotor interchanged. Because the rotor must now be driven in a a direction opposite to that of Case E, it therefore rotates in opposite direction to the rotating field of the stator, thereby establishing a slip s greater than unity.

The resulting power amplification is defined by the ratio

$$A = \frac{-\Pi_r}{\Pi_s} = s$$

Induction-Machine Systems for Constant-Frequency Generation

A variable-speed constant-frequency generating system using induction machines can be viewed from an energy-conversion standpoint as constituting essen-

tially the inverse of a variable-speed a-c motor system. Whereas the problem of speed control of a-c motors has been studied from the earliest days of a-c power systems, it is only recently that engineers are becoming interested in the use of induction machines for generation, especially in air-borne-type applications.

One general and basic method for controlling the speed of an a-c motor consists of inserting voltages of the appropriate frequency, magnitude, and phase in the rotor circuits of the motor. The principle underlying this method of induction machine operation has been discussed under Cases A and C in the previous section. It has led to the invention of numerous configurations known by such names as Kramer, Leblanc, Scherbius systems which have been applied successfully to large polyphase induction motor installations as encountered, for example, in rolling mills and wind-tunnel drives. These systems make use of auxiliary commutator-type machines which act as frequency changers to supply or recover the rotor slip-frequency power without incurring excessive ohmic losses. The Schrage brush-shift motor is another example of a machine system in which frequency changer and main motor are combined in one frame.

All these induction-motor systems can be operated in an inverse manner to supply constant-frequency power when driven from a variable-speed source. However, when considering their use for air-borne applications, the fundamental laws relating to the flow of power in the induction machine should be borne in mind as discussed under Cases B and E or F. Operation above and below synchronous speed can be realized; however, the larger the speed variations (or slip), the larger the size required for the auxiliary apparatus. Furthermore, the use of auxiliary commutator machines may present severe problems in the kind of operations and environments encountered in air vehicles. To eliminate these problems, consideration should be given to the development of new static frequency-conversion equipment employing high-power switching transistors and magnetic cores which could be used in conjunction with induction machines to produce constant-frequency generation.

A different arrangement for providing rotor slip-frequency power to an induction machine without recourse to commutator or rectifier-type apparatus makes use of another induction machine. The configuration built in a single frame may be designated as a double-induction-machine system.[8] This generating system, consisting of two series-connected, mechanically coupled rotors, may be regarded as the counterpart of the self-synchronous system comprising two uncoupled induction motors with interconnected rotor circuits. This machine system running above synchronous speed ($s<0$) is equivalent to an induction generator with variable stator-to-rotor coupling. Whereas an ordinary single-induction generator with fixed load and constant frequency must operate at a fixed slip, the double-induction generator can function with variable slips. However, since there is no way of recovering the slip-frequency rotor power without introducing some rectifier-type device, the efficiency of the double-induction generator is still directly governed by the value of slip exactly as in the case of the ordinary induction machine. Hence when large speed variations are encountered as in many air-borne systems, the efficiency of the induction generator is unavoidably low. Another basic disadvantage is that it is cannot supply lagging reactive power which must then be furnished by suitable capacitors or synchronous machines connected in parallel. The power transfer characteristics of the double-induction generator may be explained by using the block diagrams of Fig. 3. At low values of slip, the system corresponds to the combination of two blocks described by Fig. 3(B) in which both rotors supply as outputs Π_r, their total copper losses. For large slip values, the situation is that of a cascaded arrangement of Figs. 3(A) and 3(B) in which the rotor of one machine supplies power to the other rotor in addition to the total rotor copper losses. The result is that this machine will now operate as a motor above synchronous speed. There is still an over-all generator action, but the efficiency is low.

The characteristics of the double-induction generator are analyzed in some detail in the Appendix.

Conclusions

The simplicity and flexibility exhibited by the induction machine in providing electromechanical energy conversion make it attractive for use in constant-frequency a-c power generating systems. As a drive motor, the machine can be controlled to run at constant speed from a variable-frequency source; as an exciter with integral rotating rectifiers, its rotor power can be used to furnish the field excitation of a conventional alternator driven at constant-speed;[9] as a main generator, it can be combined with suitable auxiliary induction or commutator machines to produce constant-frequency a-c power when driven from a variable-speed source.

However, in considering these machine arrangements for air-borne electric systems, the electromechanical energy-conversion aspects must clearly be kept in mind. From this standpoint, the induction machine is seen to be a generalized electromechanical converter absorbing (or delivering) electric power in its input (or output) stator member, converting (or receiving) the fraction $(1-s)$ of this power as mechanical power, and transforming (or absorbing) the fraction s as electric power in its output (or input) rotor member with a corresponding change of frequency. These power-flow relations establish inherent physical limitations to the design of induction-machine systems as they constitute the governing factors which determine the magnitudes of the efficiencies and the sizes of the auxiliary equipment required, such as capacitors, commutators, or induction machines. By themselves, induction machines are basically inefficient when running at large slips. This characteristic is common to any device in which slip occurs between an input and an output member. To improve the efficiency of the induction-machine system, some form of controllable rectifiers must be used to recover or supply the slip-frequency power. Although this action is conventionally performed by commutator-type machines, switching transistors and magnetic cores offer interesting new possibilities for accomplishing the task.

In addition to these fundamental aspects of energy conversion in induction-machine systems, other problems must also be recognized and studied in order to determine the suitability of these systems for constant-frequency variable speed electric-power generation. Confining attention to purely electrical matters, the following related considerations are of particular importance in air-borne-type applications.

1. The parallel operation of generating systems.
2. The regulation of frequency, voltage, and load division.
3. The transient performance and stability of the generating system.
4. The harmonic content of the output generated voltages.

The satisfactory solution of such problems will undoubtedly contribute to the successful development of constant-frequency variable-speed a-c generators. A generating system employing the induction-machine principle appears to be one possible form of realization which, though

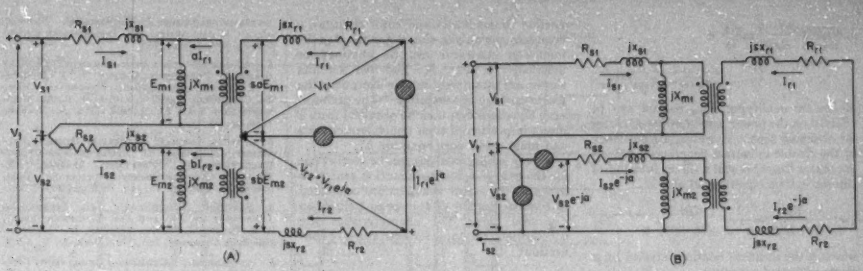


Fig. 5. Equivalent circuit of double induction machine (stator connected in series)

A—Phase shifter in rotor
B—Phase shifter in stator

beset by inherent physical limitations, may still prove to possess definite advantages in some air-borne electric systems.

Appendix. Analysis of the Double-Induction Generator

The double-induction generator consists of two mechanically coupled induction machines with their stators connected in series (or in parallel with the possible interposition of a variable ratio transformer) and their rotors in series. An appropriate regulator including a capacitor excitation system is assumed present in the stator circuits to control terminal voltage and frequency. A phase shift between rotor impressed voltages is introduced by displacing the rotor windings relative to each other while maintaining the stator windings in phase. This phase shift can also be obtained by a stator displacement with the rotors in phase or by a combination of both stator and rotor displacements. These displacements can be made either mechanically or electrically by means of additional windings, usually of the so-called saturating type.

The steady-state equations of the double-induction generator are first written assuming that the two induction machines are not connected either mechanically or electrically. These equations therefore assume the form given in equations 1 and 2 and are best expressed using matrix notation as

$$[V]=[Z]\times[I]$$

or

$$\begin{bmatrix} V_{s1} \\ V_{r1} \\ V_{s2} \\ V_{r2} \end{bmatrix} = \begin{bmatrix} R_{s1}+jX_{s1} & jaX_{m1} & & \\ jasX_{m1} & R_{r1}+jsX_{r1} & & \\ & & R_{s2}+jX_{s2} & jbX_{m2} \\ & & jbsX_{m2} & R_{r2}+jsX_{r2} \end{bmatrix} \times \begin{bmatrix} I_{s1} \\ I_{r1} \\ I_{s2} \\ I_{r2} \end{bmatrix} \tag{15}$$

The turns ratios are denoted by a for machine 1 and by b for machine 2.

The next step in the analysis is to determine the constraints established by connecting the two induction machines to form the given system configuration. These constraints reduce to equating the slips in the two machines and to writing the following relation between the old variables (unprimed) and the new variables (primed):

$$[I]=[C]\times[I']$$

or

$$\begin{bmatrix} I_{s1} \\ I_{r1} \\ I_{s2} \\ I_{r2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & -e^{j\alpha} \end{bmatrix} \times \begin{bmatrix} I_s \\ I_r \end{bmatrix} \tag{16}$$

The constraint matrix $[C]$ is established on the assumption that the stators are in phase and connected in series while the series-connected rotors have a phase shift of α. The impedance matrix $[Z']$ describing the actual system is then obtained from the equation

$$[Z']=[C_t^*]\times[Z]\times[C]$$

where C_t^* means the conjugate transpose of C. After carrying out the foregoing matrix operations, the performance equations of the double induction machine are finally written as

$$\begin{bmatrix} V_t \\ 0 \end{bmatrix} = \begin{bmatrix} R_{ss}+jX_{ss} & jaX_{ms}-jbX_{ms}e^{j\alpha} \\ jasX_{ms}-jbsX_{ms}e^{-j\alpha} & R_{rr}+jsX_{rr} \end{bmatrix} \times \begin{bmatrix} I_s \\ I_r \end{bmatrix} \tag{17}$$

where

$R_{ss}=R_{s1}+R_{s2}$; $X_{ss}=X_{s1}+X_{s2}$

$R_{rr}=R_{r1}+R_{r2}$; $X_{rr}=X_{r1}+X_{r2}$

V_t = terminal voltage $= V_{s1}+V_{s2}$

The electromagnetic torque is given by

$$T_d=\frac{p}{\omega_s}\Re e[(jaX_{ms}-jbX_{ms}e^{-j\alpha})I_s^*I_r] \tag{18}$$

It is interesting to note that the equations of the double-induction machine are similar to those of a single-induction machine given in equations 1, 2, and 3. The main effect of the double connection appears in the off-diagonal term representing the stator-rotor mutual coupling. Since this coupling may be varied by changing the turns ratio b/a and the phase shift α, it provides an added degree of flexibility which can be utilized to achieve constant-frequency variable-speed system operation. This is in contrast to the single-induction generator capable of operating at only one value of slip for any given load.

The solutions of equation 17 are

$$\frac{I_s}{V_t}=\frac{R_{rr}+jsX_{rr}}{R_{ss}R_{rr}-s(X_{ss}X_{rr}-X_{mm}^2)+j(R_{rr}X_{ss}+sR_{ss}X_{rr})} \tag{19}$$

and

$$\frac{I_r}{V_t}=\frac{bsX_{ms}\sin\alpha+j(bsX_{ms}\cos\alpha-asX_{ms})}{R_{ss}R_{rr}-s(X_{ss}X_{rr}-X_{mm}^2)+j(R_{rr}X_{ss}+sR_{ss}X_{rr})}$$

where

$$X_{mm}^2=a^2X_{ms}^2+b^2X_{ms}^2-2ab\cos\alpha X_{ms}X_{ms}$$

The stator power is therefore

$$P_s=\Re e[I_s^*V_t]=\frac{V_t^2(R_{ss}R_{rr}^2+sR_{rr}X_{mm}^2+s^2R_{ss}X_{rr}^2)}{[R_{ss}R_{rr}-s(X_{ss}X_{rr}-X_{mm}^2)]^2+(R_{rr}X_{ss}+sR_{ss}X_{rr})^2} \tag{20}$$

and the rotor power is

$$P_r=I_r^2R_r=\frac{V_t^2R_{rr}s^2X_{mm}^2}{[R_{ss}R_{rr}-s(X_{ss}X_{rr}-X_{mm}^2)]^2+(R_{rr}X_{ss}+sR_{ss}X_{rr})^2} \tag{21}$$

It is clear from equation 20 that, if s is negative, the stator power may be negative indicating the delivery of a net power out of the double-induction generator. If stator resistances are neglected, the stator power expression becomes

$$\Pi_s=\frac{V_t^2sR_{rr}X_{mm}^2}{s^2(X_{ss}X_{rr}-X_{mm}^2)^2+X_{ss}^2R_{rr}^2} \tag{22}$$

and the power factor angle

$$\theta_s = \tan^{-1}\frac{sX_{rr}}{R_{rr}} - \tan^{-1}\times$$

$$\frac{R_{rr}}{-s[X_{rr}-(X_{mm}^2/X_{ss})]}$$

Hence by modifying X_{mm} control can be effected on the power output or power factor for changing slips. However, the efficiency of the double-induction generator (neglecting stator losses) is still a direct function of slip since, from equations 21 and 22,

$$\eta = \frac{-\Pi_g}{-\Pi_g + I_r^2 R_r} = \frac{1}{1-s} \quad \begin{matrix}(s<0)\\(R_s=0)\end{matrix} \qquad (23)$$

which is the identical relation satisfied by a single-induction generator.

Another procedure for deriving the performance equations of the double-induction generator can also be developed with the use of the equivalent circuit shown in Fig. 2. By combining two such equivalent circuits so as to satisfy the constraints of interconnection (series stator and rotor circuits), a complete equivalent circuit for the double-induction generator can be obtained as illustrated in Fig. 5. The two circuits shown are completely similar except for the location of the phase shifter. The performance equations can then be obtained from a direct inspection of these equivalent circuits and must check with equation 17.

A very similar analysis can be carried out for the case of a double-induction generator in which the stator windings are connected in parallel instead of the series connection discussed here. The characteristics are substantially the same for both types of connections.

References

1. Engineering Design Study for a Two Unit (Electrical) Constant Speed 85 Horsepower Alternator Drive, J. M. Pestarini. *Technical Report 55-236*, Wright Air Development Center, Dayton, Ohio, June 1955.

2. Aircraft Motor Generator with Secondary Standard Frequency Output, L. J. Johnson, S. E. Rauch. *Transactions*, Professional Group on Component Parts, Institute of Radio Engineers, New York, N. Y., vol. CP-5, no. 1, Mar. 1958, pp. 28-31.

3. Future Aircraft Electrical Systems Predicted (abridgement of Direct Engine-Driven Unconventional Electric Systems), K. Martinez. *Journal*, Society of Automotive Engineers, New York, N. Y., Jan. 1958, pp. 68-70.

4. Electric Machinery: An Integrated Treatment of A-C and D-C Machines (book), A. E. Fitzgerald, Charles Kingsley, Jr. McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

5. Alternating Current Generating Arrangement for Constant Frequency, K. Polasek, K. A. Jousson. *U. S. Patent no. 2,747,107*, May 22, 1956. Also, *British Patent no. 700,036*, Sept. 18, 1951.

6. Arrangement Allowing the Elimination of Brushes in Electrical Machinery, M. E. Rémy. *Revue Générale de l'Electricité*, Paris, France, vol. 64, no. 3, Mar. 1955, pp. 113-18.

Communication Systems for Railway Traffic Control

H. C. SIBLEY
MEMBER AIEE

CONTROL of railway traffic from a central location requires the transfer of large amounts of information between the central location and many points along the railroad. The number of wayside information points depends on the type of signaling installed, which in turn depends on the type and amount of traffic handled by the railroad. Certainly, effective communication is required with all locations where track switches and controlled signals are situated. Information must also be available from a sufficient number of points to keep the dispatcher informed about the location and progress of all trains.

Additional information of a specialized nature is rapidly becoming a significant part of traffic control. An example of this special information is the signal picked up by hot box detectors. Voltages proportional to the temperature of each journal box are transmitted to the dispatcher's office as freight trains pass remote detector locations at normal speed.

The General Railway Signal Company designs and builds many communication systems for railroad signaling. Three of these systems will be described in this paper.

Centralized Traffic Control

Centralized traffic control (CTC) requires reliable 2-way communication between a control office and a number of wayside locations. A typical application of CTC may extend over 200 miles with 40 or more field locations. The plant at a field location may comprise a track switch and signals at the end of a passing siding or may be a fair sized interlocking, that is, a group of switches and signals such as may be found at a terminal or at a junction, previously controlled by a local operator in a tower. A CTC system enables the dispatcher to operate track switches and signals at any location. It also provides information as to the position of all wayside apparatus and the location of trains.

Control System

The control system is the part of a CTC system which provides communication outbound from the control office to any of the wayside locations. The information input to this system is supplied by levers or pushbuttons operated by the dispatcher to control track switches or signals. These levers or pushbuttons are generally located on a control machine or console, placing an entire division of a railroad virtually at the dispatcher's fingertips. A typical control machine is shown in Fig. 1. The output of the control system positions relays at the field locations.

One type of CTC control is a synchronous system using relays and transmitting a polar code. In this system a control code comprises 15 equal time intervals of direct voltage line energization of either polarity. When the system is at rest the line is held energized with negative polarity. Mechanical oscillators provide the time base for the system.

Fig. 1. Typical control machine

Paper 59-249, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 23, 1958; made available for printing December 10, 1958.

H. C. Sibley is with General Railway Signal Company, Rochester, N. Y.

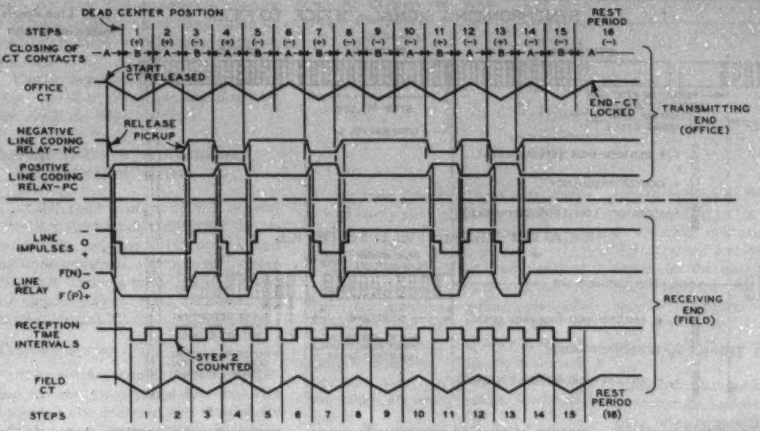


Fig. 2. Sequence of line events, control system

They are mechanically tuned to a frequency of $5^1/_2$ cycles per second and operate counting relays every half cycle. At the start of a control cycle, the oscillators at the control office and at all field locations are started. A binary counter comprising 4 relays counts the 15 steps of the code and the period of rest. The duration of a control cycle is approximately $1^1/_2$ seconds.

Fig. 2 shows the sequence of line events of a control cycle. The top two lines are the contact and armature operations of the mechanical oscillator at the control office (designated OFFICE CT). The coding relay operations are shown by the next two lines. The line impulses are the result of the closing of the coding relay contacts.

At the field locations the line relay closes its *P* or *N* contacts, depending on the polarity of line pulses. The mechanical oscillator (FIELD CT) defines the beginning and end of each step in the field. The line labeled RECEPTION TIME INTERVALS shows the portion of each step available for station selection or function control. These intervals are each equal to the duration of a step less the operating time of the counting relays.

Fig. 3 shows, in block form, the components of an office control unit and their functional relationship. The line filter is required to remove the higher harmonics of the code to prevent interference with voice or carrier circuits on the same pair of line wires.

To transmit a control cycle, the dispatcher positions the control levers to set up a route and then presses a start button. One start button is provided for each field location. The appropriate station selection relay is energized to cause the coding relays to generate the proper station selection code. After station selection, the control levers determine the code transmitted for the remainder of the cycle. The oscillator establishes the time of operation of the coding relays. The counting relays serve as a commutator to switch the control of the coding relays among the several control levers and the station selection relay.

Fig. 4 shows a field control unit. The line relay, when it operates for the first time in a control cycle, starts the oscillator. For the remainder of the cycle, the oscillator operates on stored mechanical energy and is not affected by external forces until 15 steps have been counted. Then, the office control unit enters a period of rest, placing continuous energy on the line, and all mechanical oscillators are restored to their latched or rest position. The line relay detects the polarity of line energization.

The function relays are controlled by the line relay and counting relays after the location has been selected by the first part of the control cycle.

The local energy supply for the control system is 28 volts furnished by battery or a rectifier. The line energy for d-c polar operation is a rectifier. The voltage depends on the length and type

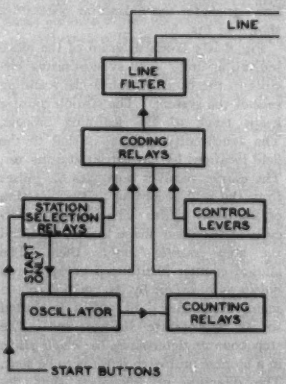


Fig. 3. Office control equipment

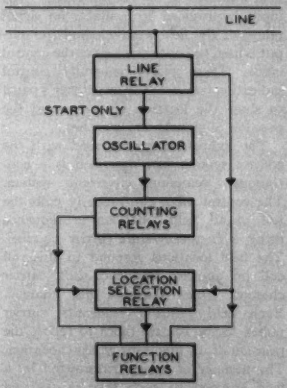


Fig. 4. Field control equipment

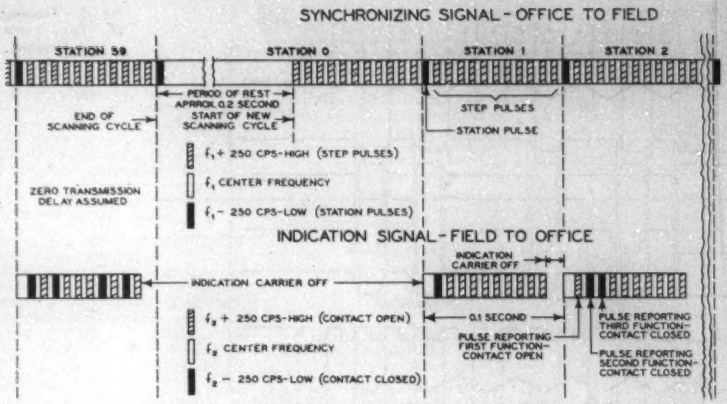

SYNCHRONIZING SIGNAL - OFFICE TO FIELD
STATION 59
STATION 0
STATION 1
STATION 2
PERIOD OF REST - APPROX. 0.2 SECOND
END OF SCANNING CYCLE
START OF NEW SCANNING CYCLE
STEP PULSES
STATION PULSE
f_1 + 250 CPS-HIGH (STEP PULSES)
f_1 CENTER FREQUENCY
ZERO TRANSMISSION DELAY ASSUMED
f_1 - 250 CPS-LOW (STATION PULSES)
INDICATION SIGNAL - FIELD TO OFFICE
INDICATION CARRIER OFF
INDICATION CARRIER OFF
f_2 + 250 CPS-HIGH (CONTACT OPEN)
0.1 SECOND
PULSE REPORTING THIRD FUNCTION - CONTACT CLOSED
f_2 CENTER FREQUENCY
PULSE REPORTING FIRST FUNCTION - CONTACT OPEN
PULSE REPORTING SECOND FUNCTION - CONTACT CLOSED
f_2 - 250 CPS-LOW (CONTACT CLOSED)

Fig. 5. Line signals, indication system

of line and the number of field locations.

A code repeater is used when the line length and number of field locations exceed the capacity of one line power supply or produce excessive distortion of code pulses. The repeater reshapes the code pulses and keys a line power supply at the repeater location.

A carrier system employing two-way frequency-shift keying is provided when the system is connected to a line already used for a d-c service. The frequency of the carrier transmitter is shifted 100 cycles per second above or below the nominal frequency to provide the carrier equivalent of a polar code.

Indication System

The indication system provides communication inbound from the many field locations to the control office. The input is provided by contacts on signal relays at the field locations and the output is used to position relays at the control office. These relays generally control indication lamps on the dispatcher's panel to show the location of trains and the position of signals and switches.

The indication system used with the control system just described is a continuously scanning, electronic system. The control office continuously calls the roll of all field locations in sequence using a frequency-shift carrier channel. The field locations respond to this roll call on another frequency-shift carrier channel, time sharing this channel. Each field location transmits carrier pulses to the control office to report the position of all functions at that location. The number of pulses transmitted by a field location on each scan is assigned in multiples of ten depending on the amount of wayside apparatus to be reported by that location. Each group of ten pulses is called an indication station. Each pulse represents a two-position field function. If a certain pulse represented a signal, it would be transmitted at one edge of the carrier channel if the signal indicated "stop," or at the other edge if the signal indicated "proceed." In general, an open indication contact results in a pulse of high frequency, that is, center frequency plus 250 cycles per second. A closed indication contact results in a pulse of low frequency, center frequency minus 250 cycles per second. The scanning rate is 100 indications per second. The length of a scan, and therefore the repetition rate, depends upon the number of field stations. A typical installation reports each of 600 field functions every 6 seconds. The line signals for the indication system are shown in Fig. 5.

Fig. 6 is a block diagram of the office indication unit. The synchronizing circuits in the timer establish the scanning rate of the system. The station counter keeps track of the scanning process. The synchronizing signal is sent to the field locations by the carrier transmitter. The carrier receiver separates the pulses from the field according to the direction of frequency shift. The indication circuits in the timer shape these pulses and the step counter counts them. The station counter identifies a group of ten pulses as to origin by its position in the scan. Each group of pulses is routed to its appropriate storage unit. The step counter determines to which stage in a storage unit each pulse is to be sent. The direction of frequency shift determines whether or not the relay connected to the selected stage of the storage unit is to be energized.

Fig. 7 shows the organization of the field indication unit. The carrier receiver feeds the synchronizing signal to the scanner where it is shaped to drive the station counter. Jumpers on the station selector determine at which count the indication unit will transmit its signal to the control office. The indication contacts and step counter work through the diode matrix to form the indication signal.

Cold cathode tubes are used for pulse counting and storage. Considerable energy is saved by using these tubes since they have no heaters. The fact that cold cathode tubes glow during conduction enables the maintainer to make routine checks and locate trouble

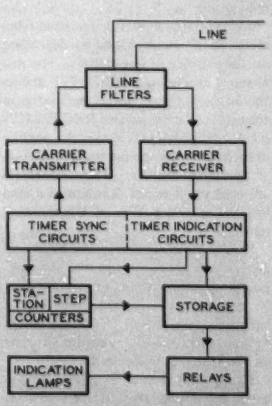


Fig. 6. Office indication equipment

with no more instrumentation than a voltmeter. Operation of these tubes at well below rated current has resulted in long life.

Carrier repeaters of the regenerative type as well as the simple amplifier type are available when required.

Special Features of Centralized Traffic Control System

At this point it might be well to mention the fundamental difference in requirements between CTC control and indication systems. Both systems have electrical contacts as inputs and relays as output devices. The differences are in the application.

Control System

The control system has inputs originating at one geographical location with outputs to be delivered to a multiplicity of locations. Further, the inputs originate in the mind of the dispatcher and can occur no more rapidly than the time required for him to plan train movements and express his wishes by operating control levers or pushbuttons.

The dispatcher executes his plans at one location at a time. The speed required of a control system is that necessary to deliver controls to each field location as quickly as the dispatcher can set them up. The reliability required of a control system is the highest attainable because the apparatus at a field location is not capable of human judgment and will attempt to execute any control called for. An error in transmission, if undetected, could set up a wrong route or no route at all,

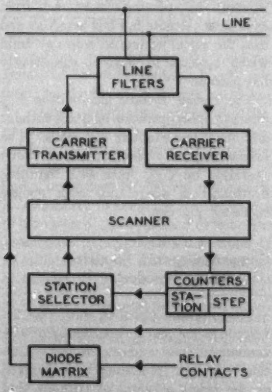
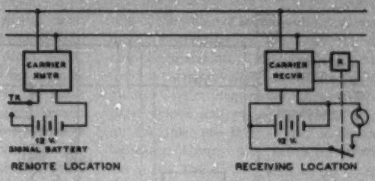


Fig. 7. Field indication equipment

Fig. 8. Diagram of tone system arrangement for one indication



causing train delays. There is no question of safety since all safety circuits are local and no control resulting in an unsafe condition can be executed.

The conditions outlined as to speed and reliability requirements are well met by the relay stepping system previously described. The relays used are designed for extremely long life and reliable contact operation. The control code includes a number of checks reducing the possibility of wrong routes essentially to zero.

Indication System

The indication system has inputs originating anywhere along a railroad at random times. If there are a dozen trains operating on a division, the information regarding train locations can change frequently in a few seconds as several trains simultaneously cross short track circuits. The electronic scanning system provides the speed and capacity to satisfy these conditions.

Multistation Operation

CTC systems differ from conventional communication systems in that they are multistation systems rather than end-to-end. This fact means that every relay or tube added to a field unit is multiplied by 25 or more on any typical installation. It is, therefore, very desirable to keep these units as simple as possible.

Furthermore, since there are so many wayside locations, the line apparatus must be designed to bridge rather than terminate the line and present as little bridging loss as possible.

Since the field locations are unattended, the carrier system must be capable of operating for long periods without adjustment. The carrier receivers are designed to work over a wide range of input signal level. At the time of installation, transmitter output and receiver sensitivity are adjusted for average line conditions. The tolerance in the receiver circuits will allow for variations in line attenuation and tube aging in the carrier units.

Check circuits to prevent a failure in a field location from interfering with the proper operation of the rest of the system are essential. Electronic apparatus common to the entire system, such as the office indication unit and carrier repeaters, are installed in duplicate. A switch on the dispatcher's control machine transfers the operation to the standby equipment in case of trouble.

Block Occupancy Reporting

Railroads having high traffic densities, for example, subways and commuter lines, often require the reporting of train movements over many short track circuits or blocks. A communication system is required which is capable of transmitting many digits of information with only one or two digits originating at any one location. Obviously, this information can be handled by multiconductor cable with d-c or power frequency. To conserve line wire some type of multiplexing is desirable.

A simple tone system is one means of doing the job. Tone transmitters of different frequencies are situated at the various blocks and are keyed when these blocks are occupied. Tone receivers at the dispatcher's office operate relays to energize lamps on a miniature track diagram. Fig. 8 shows this type of system.

An extension of the tone system provides two independent digital indications on one tone channel. The pair of line wires over which the tone system operates carries a direct voltage which is pole-changed 75 times per minute. During the time that the line voltage is of one polarity, each tone channel carries one indication. When the line voltage is switched to the opposite polarity, each tone channel carries a different indication.

The tone transmitters and receivers are transistor operated. The transmitters develop +8 dbm (power in decibels referred to 1 milliwatt) output with 0.020 ampere input at 12 volts. The receivers draw 0.005 ampere with no signal and 0.040 ampere when receiving a signal and operating a relay.

Fig. 9 shows the basic circuit arrangement of the system. Relay CT oscillates continuously applying direct voltage to line wires *LW*1 and *LW*2, alternating

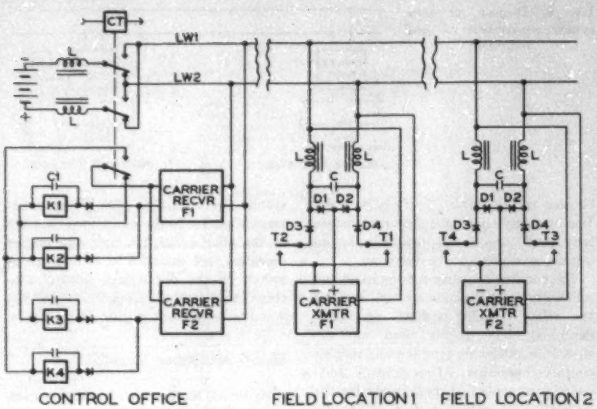


Fig. 9. Diagram of block occupancy reporting system

the polarity each time its contacts change position. A contact of relay CT simultaneously changes the output of carrier receiver *F*1 between relays *K*1 and *K*2 and the output of receiver *F*2 between relays *K*3 and *K*4.

Carrier transmitter *F*1 has its energy input terminals connected to the line through the two relay contacts *T*1 and *T*2, the four diodes and the two inductors marked *L*. The carrier output circuit of the transmitter is connected to the line above the inductors. The purpose of the inductors is to isolate the energy input terminals of the transmitter from the line at carrier frequencies.

The capacitor in parallel with diodes *D*1 and *D*2 removes the peaks from the pole-changed line voltage.

If relay contacts *T*1 and *T*2 at field location 1 are closed, the four diodes form a full-wave bridge rectifier. The path through this bridge from line wire *LW*1 to the minus terminal of the transmitter passes through diode *D*3 and contact *T*2. The path from *LW*2 to minus passes through diode *D*4 and contact *T*1. The return path from the plus terminal of the transmitter to the opposite line wire is through diodes *D*2 and *D*1 respectively. Therefore, contact *T*1 controls the transmitter for one line polarity and *T*2 controls it for the opposite polarity. Further examination of the circuit will show that relay *K*1 is connected to the output of receiver *F*1 at the time that transmitter *F*1 is controlled by contact *T*1. Similarly, relay *K*2 is connected to the same receiver when contact *T*2 is controlling the transmitter.

The capacitors across the relay coils hold the relays energized during alternate swings of relay CT. The inductors *L* at the control office prevent the line battery from loading the line at carrier frequencies.

Several carrier transmitters may be connected to the line at one location using one line coupling unit, the assembly including the inductors, capacitor and part of the diode bridge. The length of line and total number of transmitters which may be used is limited by the voltage and current rating of the line battery or power supply. A recent installation uses 28 carrier transmitters distributed along 11 miles of 14-gage twisted pair in underground duct. Thirty-seven channels are available between 430 and 7,300 cycles per second.

The line circuit and energy input circuit of the transmitters and receivers are designed to be protected by standard railway signal-type lightning arresters (1,000-volt breakdown).

A unique feature of this system is the fact that local energy is not required for the transmitters at the field locations. This arrangement is ideal for a system of this type which often derives inputs from contacts of track relays at locations where a 12-volt battery is not available.

A 12-volt d-c supply is required at the control office to energize the carrier receivers.

Analog Data Transmission

The trend in railroad operation is toward control of longer lengths of territory from one central location. As consolidation of control occurs, hundreds of miles of railroad are operated with no personnel located directly at the wayside. The two systems just described provide a means to control switches and signals and to receive information concerning train movements over these distances. Sometimes it is necessary to relay analog information from the field to the control office. This information may be any signal which is variable over a continuous range such as the level of fuel in storage tanks, the water level of a river at flood stage, or pressures and temperatures in some device.

The analog data system comprises a single-side-band carrier transmitter and receiver with submodulation equipment for frequency-modulated telemetering. The analog data, in the form of voltages, vary the frequency of subcarriers by amounts proportional to the voltages. At the receiving location, the variable frequency subcarriers are converted to voltages identical to the input signals. Subcarrier equipment is available for two or four analog channels. Two digital channels may be provided for additional on/off indications, timing, or marker signals.

An interesting application of this system is in conjunction with a hot-box detector. The hot-box detector is an infrared-sensitive device situated close to the rails. The detector actuates a pen recorder to indicate the relative temperatures of freight car journals as trains pass by at normal speed. An analog data transmission system makes it unnecessary for an attendant to check the record in the field each time a train goes by. The recorders may be located in the dispatcher's office. Where CTC is used, the operator checks each train as it passes remote hot-box detectors and uses the signal system to stop any train which appears to have a dangerously hot journal.

The carrier system has built-in line filters to allow operation on a line without additional line coupling apparatus. The insulation level of the line circuit allows operation on CTC lines with lightning protection in accordance with regular signaling practice.

Compatibility with Standard Signaling Practice

The railway signaling art has developed in the past independently of wire communications and electronics. Certain standard practices have evolved as the result of years of rigorous field experience. Continuity of service and

maintainability are important requisites of railroad signal equipment.

One important practice is the use of 3,000-volt insulation in line and power circuits. This insulation level permits the use of lightning arresters with sufficiently large gaps to make it unlikely that a discharge will result in a permanent short circuit. Transistor operated apparatus requires voltage limiting devices in addition to lightning arresters to protect them from voltage surges that are tolerated by conventional signal apparatus.

Another practice which must be considered is the method of wiring apparatus housings. This practice has developed with d-c and power circuits in which wire length and arrangement are not important to circuit performance. Relay housings are generally wired with no. *16* wire with very heavy insulation arranged in laced cables and connected to rugged terminals. Contrast this wiring with electronic practice with its small-gage wire, small terminals, and often short, carefully dressed leads. Electronic chassis used in signaling must allow the same external wiring practice as other devices such as relays. All external leads must be tolerant of exposure to voltage surges. The impedance of external circuits must be low and pulse widths and rise times must be great enough so that deterioration due to wiring capacitance is not serious.

The apparatus situated along the wayside must withstand vibration and severe humidity and temperature. Railroad specifications require testing over an ambient temperature range of −40 to 160 degrees Fahrenheit. Most installations are unattended and therefore must not require frequent adjustment. The carrier systems previously described will operate with variations in line attenuation of the order of 20 db without adjustment. Adjustments are kept to an absolute minimum. Level and sensitivity controls are provided for carrier equipment. Carrier frequencies, pulse widths, pulse delays and pulse repetition rates, for example, are fixed. Test jacks and indicator lamps are provided where they will help the maintainer. All chassis are provided with plug connectors so that a defective unit may be replaced quickly.

In general, electronic equipment for railway signaling use has been designed to be compatible with long established signal practice and to take its place on the railroad alongside of signal relays and other standard signal devices.

Reference

1. SYNCROSCAN—A HIGH-SPEED CONTROL SYSTEM FOR RAILWAY SIGNALING, H. C. Sibley. *Electrical Engineering*, vol. 76, Jan. 1957, pp. 38–43.

Rapid Heating of Dielectric Materials at 915 Mc

G. E. FEIKER
ASSOCIATE MEMBER AIEE

N. C. GITTINGER
NONMEMBER AIEE

THE HEATING of dielectric materials by means of microwave energy was investigated in the General Electric Company a decade ago by members of the Research Laboratory, who demonstrated the feasibility of preheating plastics and heating foods using a magnetron oscillator at a frequency of 1,000 mc (megacycles). In recent years the *GL-6787* magnetron, operating in the 915-mc Industrial-Scientific-Medical band, has been developed as a source of power for these applications. In quantity production, this tube will provide economical microwave power for consumer and industrial applications. In order to design equipment for efficient preheating, it is necessary to know the characteristics of the materials as a function of temperature. This paper presents the results of an initial investigation of the effectiveness of 915-mc heating on various dielectric materials, and describes means of measuring the characteristics of materials at this frequency.

Paper 59-30, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing November 19, 1958.

G. E. FEIKER and N. C. GITTINGER are with the General Electric Company, Schenectady, N. Y.

Basic Considerations

The power absorbed by a dielectric subjected to an alternating electric field can be calculated from the equation:[1]

$$P = 0.55 f \epsilon'' E^2$$

where

P = the power absorbed, watts/centimeter^3
f = the frequency, mc/second
E = the rms voltage gradient, kilovolts/centimeter
ϵ'' = the loss factor of the material

The rate at which a material can be heated by dielectric heating depends on two factors:

1. The loss factor ϵ'' of the dielectric. This is the property of the material which gives the loss per unit volume for a given voltage gradient.
2. The frequency of operation.

The gradient E is usually limited by the breakdown of air between electrodes and the dielectric.

Since the loss factor is very low for many important dielectrics, the frequency of operation should be made as high as practicable in these applications. Because presently available dielectric heating apparatus operates at frequencies below 100 mc, the application of 915 mc makes it possible to heat some materials which previously could not be heated by this means.

Description of Experimental Equipment

The equipment set up for the tests to be described is shown in Fig. 1. It consists of a power supply and magnetron oscillator, directional couplers, slotted lines, heating cavity, and associated instrumentation. The oscillator employs the *GL-6787* metal and ceramic magnetron, having a rated continuous-wave output of 2.5 kw. The power supply associated with the tube is capable of operating the tube at any power level from 50 watts to the full 2.5 kw. The magnetron output is coupled to a system of 3$^1/_8$-inch-diameter coaxial line, which feeds the heating cavity. This cavity consists of a cylinder with re-entrant posts, shown in Fig. 2. The sample to be heated is placed in the air gap between the two re-entrant posts, where it is subjected to an intense electric field when the cavity is resonated. The cavity is tuned by adjustment of the gap between the re-entrant posts and is coupled to the line by means of an adjustable coupling loop in one side of the cavity. By properly ad-

Fig. 1. 915-mc dielectric heating equipment with manually tuned heating cavity

Fig. 2. Interior view of heating cavity showing coaxial thermocouple inserted in sample

justing the coupling loop and the tuning gap, this cavity will match a wide variety of materials to the 50-ohm coaxial transmission line.

Measurement equipment in the coaxial line consists of a dual directional coupler arrangement for continuous measurement of power and VSWR (voltage standing wave radio), and a slotted line section for impedance measurements. Temperature measurement of the dielectric samples in the cavities is accomplished by means of a special thermocouple. The thermocouple element is of coaxial construction, and is enclosed in a stainless steel shielding tube over its entire length, except at the actual junction. The shielding tube prevents thermocouple self-heating caused by the strong electric field within the cavity. The thermocouple is installed from the side of the cavity perpendicular to the electric field, so as to reduce heating of the thermocouple shield as much as possible. The directional couplers and thermocouple are connected to photorecorders. This arrangement permits measurements of these quantities during rapid heating tests.

Comparative Tests at 915 Mc and 70 Mc

It has been stated previously that the rate of heating of a given material is proportional to frequency. As a consequence, it should be possible to heat some materials at 915 mc which are difficult or impossible to heat in conventional preheaters operating at frequencies below 100 mc. In order to investigate this possibility, comparative tests of heating effectiveness were made at 915 mc and 70 mc, using a 3.5-kw preheater manufactured by W. T. LaRose and Associates, Inc. A variety of materials was obtained for these tests as shown in Table I. Samples were heated in both machines under similar conditions, and from the results, the following conclusions were drawn:

1. If the loss factor of a material exceeded 0.01, the material could be heated at either frequency. The 915-mc equipment could heat it faster for a given voltage gradient, however.

2. If the loss factor of the material was between 0.01 and 0.001, the low-frequency heater heated the material very slowly. The 915-mc equipment could heat these materials at a relatively high rate.

3. If the loss factor of the material was below 0.001, neither equipment could heat it well.

4. In most cases, the limitation of heating in the lower frequency unit occurred because of arcing. At 915 mc, arcing rarely occurred at the powers used in the tests (100–600 watts).

Materials such as Rubber Styrene, Teflon *100X*, Fluorothene, and Styrene Acrylonitrile could not be heated well at the lower frequency, but were heated easily at 915 mc. A special low-loss glass was also heated to the melting point at 915 mc, but could not be heated at the lower frequency. Final temperature of the glass reached 800 C (degrees centigrade) at 915 mc.

Table I. Heating Tests of Various Materials at 70 Mc and 915 Mc

Material	Dielectric Constant	Power Factor	Loss Factor	Heated At 70 Mc	Heated At 915 Mc
Asbestos-filled melamine	5.2	0.035	0.18	Yes	Yes
Bakelite	4.6	0.03	0.14	Yes	Yes
Mineral-filled silicone	4.1	0.0055	0.03	Yes	Yes
Low-loss phenolic	4.3	0.01	0.04	Yes	Yes
Nylon	3.4	0.04	0.136	Yes	Yes
Lucite	3.6	0.025	0.09	Yes	Yes
Polyvinyl chloride	3.4	0.06	0.20	Yes	Yes
Fluorothene	2.3	0.0085	0.02	marginal	Yes
Styrene-acrylonitrile	2.8	0.0083	0.023	marginal	Yes
Rubber styrene	2.5	0.001	0.0025	No	Yes
Mycalex	7.4	0.0018	0.013	No	Yes
Teflon 100X	2.0	0.0005	0.001	No	marginal
Teflon	2.0	0.0001	0.00025	No	No
Polyethylene	2.3	0.0005	0.001	No	No
Polystyrene	2.5	0.0001	0.00025	No	No

Detailed Measurements on Phenolics

A detailed investigation has been made of the feasibility of rapid preheating of a general purpose phenolic compound, General Electric *12900*. As supplied for these tests, the phenolic was in the form of cylindrical preforms, $1^1/_2$ inches in diameter and 3/4 inch high, weighing 22 grams. The material is exothermic in nature.

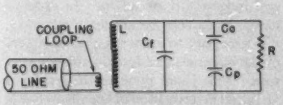


Fig. 3. Equivalent circuit of re-entrant heating cavity

L = cavity inductance
C_f = fringe capacity of electrodes
C_s = capacity of air gap between the preform and the upper electrode
C_p = capacity of preform
R = shunt resistance due to preform losses

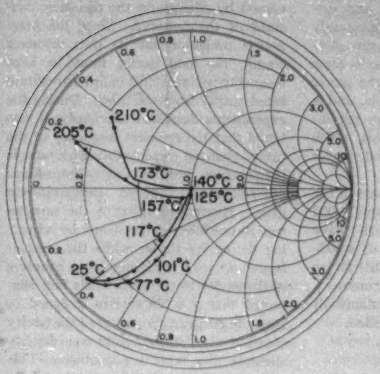


Fig. 4 (left). Smith chart of cavity impedance versus sample temperature

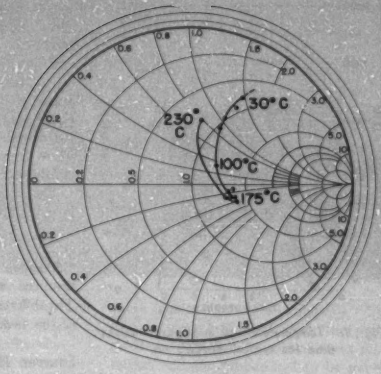


Fig. 5 (right). Smith chart of cavity impedance versus sample temperature

When heated to about 175 C. the material undergoes a chemical reaction, gets soft and pliable, and produces heat. After completion of the reaction, the material hardens into a brittle substance.

Measurements on samples of the phenolic at 30 mc showed a dielectric constant of 2.8 and a power factor of 0.06. It was felt necessary to measure these quantities at 915 mc, however, since characteristics depend on frequency in many substances. The cavity shown in Fig. 2 was used for this measurement. With a phenolic preform in place between the re-entrant electrodes, the cavity has an equivalent circuit closely approximated by the diagram of Fig. 3. In this circuit, the capacitances C_s and C_p could be calculated, whereas C_f and L were obtained by measuring samples of known dielectric constant, such as Teflon, Polystyrene, and air. By comparing the gap setting required to resonate the cavity loaded with phenolic to the settings required with the standard samples, and correcting for the knowu fringe capacitance, the dielectric constant of the phenolic could be measured.

Power factor of the phenolic was obtained by measuring the Q of the loaded cavity. This measurement determined the value of the shunt resistance R. As the unloaded cavity has a very high Q, losses in the cavity walls could be neglected, and the equivalent resistance R was assumed to be entirely attributable to the sample losses. By transposing R to a resistance in series with C_p, the power factor of the material was obtained.

Measurements made at 915 mc did not agree with those made at 30 mc. In order to check the 915-mc measurements, tests were made at several other frequencies. The results are shown in Table II.

It may be seen that the power factor and dielectric constant of the phenolic sample change considerably with frequency.

The dielectric constant and power factor of the material also change with temperature. As the phenolic heats up, there is a definite change in impedance presented to the 50-ohm coaxial line by the cavity. The change in cavity impedance with temperature, obtained by point-to-point measurements with the slotted line, is shown on the Smith charts of Figs. 4 and 5. These data were taken under different conditions of cavity tuning. They show that there is a definite problem in matching the cavity to the line, since a VSWR much above 3.0 will cause a large power loss due to reflection.

It has been found that most of the change in the sample as it heats is in the dielectric constant, which tends to detune the cavity. Heating tests have been conducted with the cavity tuned manually by adjusting gap spacing. A plot of gap spacing required for cavity resonance with temperature is shown in Fig. 6. The change in dielectric constant with temperature is found to be from a value of 2.8 at room temperature to 3.6 at 175 C.

The change in dielectric constant causes the power actually absorbed by the sample to change with temperature. Fig. 7 shows a plot of power entering the phenolic versus time for a typical heating test where

Table II. Electrical Properties of Phenolic Compound GE 12900 at Various Frequencies

Frequency, Mc	Dielectric Constant	Power Factor	Loss Factor
30	2.8	0.06	0.17
915	2.8	0.03	0.08
3050	3.2	0.04	0.13
9715	3.0	0.035	0.11

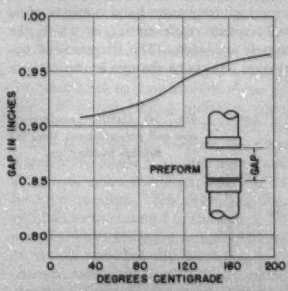


Fig. 6. Gap spacing required for resonance versus temperature, large re-entrant cavity loaded with phenolic preform

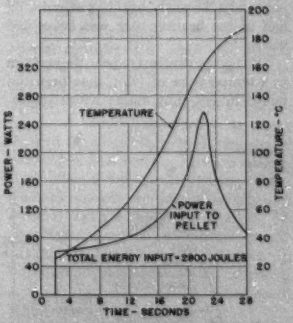


Fig. 7. Temperature rise and power input versus time for typical heating run of sample 22-gram phenolic pellet

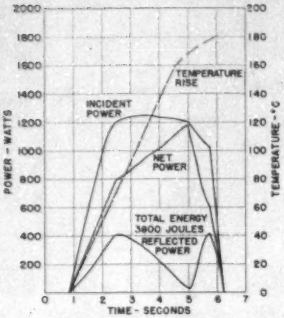


Fig. 8. Temperature rise and power versus time for typical heating run

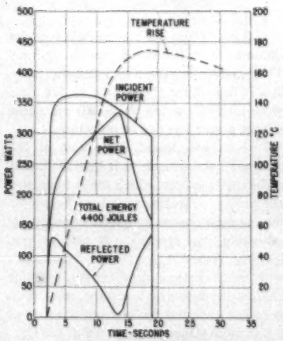


Fig. 9. Temperature rise and power versus time for typical heating run

the cavity gap adjustment was fixed. In this particular test, the power into the sample pellet at the instant of cavity resonance was 256 watts, but the power during half of the heating time was below 100 watts.

The energy required to heat the preform to a temperature of 175 C has been obtained in test runs during which the incident power, reflected power, and preform temperature were simultaneously recorded. Three typical test run plots are shown in Figs. 8, 9, and 10. In these plots, the net power input is the power actually absorbed by the sample, and is equal to the difference between the incident and reflected power curves, which were obtained from the directional coupler readings.

The average energy requirement of the large number of tests made was 4,000 joules. Theoretical calculations have shown an energy requirement of 4,300–5,000 joules, which is considerably larger than the value obtained by measurement. The difference is believed to be the result of the exothermic chemical reaction.

Uneven Heating Effects

During most of the tests, definite unevenness of heating was observed in the phenolic preforms. The interior and top of the preform were well heated, but the sides and bottom were somewhat cooler. This uneven heating was traced to two effects:

1. The uneven heating at the sides of the preform was caused by the lack of a constant field intensity at the edges of the electrodes. By using considerably larger electrodes, the sides of the preform were evenly heated.

2. The uneven heating at the bottom of the preform was caused mainly by the presence of the cold-metal electrode in contact with the preform. By suspending the preform in space between the electrodes by means of a thermocouple lead, the cold area on the bottom of the preform was eliminated. Since this method is not convenient or practical, a feasible alternative is to use internal heaters for heating the electrode to the desired final temperature.

Considerations of Preheater Design

The usual preheater operating in the very-high-frequency band consists of a self-resonant tank circuit in which the sample is placed. The frequency of operation adjusts to changes in capacitance caused by changes in the dielectric constant of the material. Since the magnetron oscillator has its own resonant circuit, the load cavity must be kept resonably well matched in order to obtain rated output from the tube. During the experiments described previously, it was found that the resonant frequency of the cavity changed considerably with sample temperature, but the cavity impedance at resonance changed comparatively little. Heating of most of the samples was successfully accomplished by adjusting the cavity tuning while the samples heated up, with no change in the coupling of the cavity to the line. This fact suggested that a servo system designed to keep the resonant frequency of the cavity constant at all times would considerably alleviate the matching problem. This servo system was built and successfully demonstrated. Its design will now be described.

The tuning servo consists of a phase detector in the coaxial line, an amplifier, and a servomotor which tunes the cavity. A diagram of the system is shown in Fig. 11.

The principle of operation of the servo-tuning system is based on the change of phase that occurs in the coaxial line as the cavity is tuned through resonance. The cavity will appear to the line as a parallel resonant circuit across the end of the line.

The cavity impedance will appear as a pure resistance only when it is tuned exactly to the operating frequency. When detuned appreciably, it will present an almost pure reactance to the line which is either inductive or capacitance, depending on which side of resonance the cavity is tuned. A large standing wave is therefore present in the line. The minimum position of this standing wave will be found to shift along the line by a half wavelength as the cavity is tuned from one side of resonance to the other side. At resonance, the standing wave minimum will be exactly half way between the positions it occupies when the cavity is detuned each side of resonance. The transition of the minimum is quite sharp, as indicated in Fig. 12(A).

This figure shows the change in minimum position which will be obtained in a

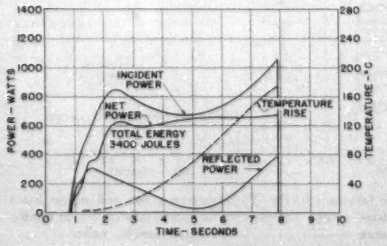


Fig. 10 (left). Temperature rise and power versus time for typical heating run

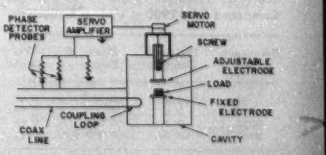


Fig. 11 (right). Schematic diagram of servo-tuned heating cavity

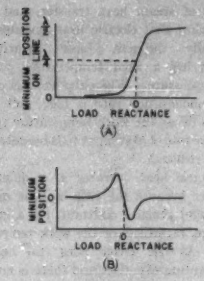


Fig. 12. Voltage minimum shift versus load reactance

A—Cavity resistance greater than characteristic impedance

B—Cavity resistance less than characteristic impedance

Fig. 13. 915-mc dielectric - heating equipment with servo-tuned heating cavity

coaxial line feeding a cavity with losses, if the resonant resistance of the cavity is greater than the characteristic impedance of the line.

The phase detector employed uses two probes in the coaxial line spaced one-quarter wavelength apart.

The probes feed a matched pair of crystal detectors with opposite polarities. The outputs of the detectors are added and fed to the servo amplifier. The polarity of the output of the crystal pair depends on which crystal is in the stronger field. If the probes straddle a voltage minimum, the output is zero, since both crystals are in the same intensity field. If the cavity is detuned, however, the standing-wave ratio will shift, and a voltage unbalance will appear at the servo-amplifier input, with polarity dependent on the direction by which the cavity is detuned. The servo amplifier feeds a servo motor which corrects the tuning.

It might seem that if the cavity is detuned sufficiently to appear as a pure reactance, the probes will straddle a maximum and no correction voltage will appear. However, the losses always present will prevent the cavity from ever reaching this condition, no matter how far it is detuned. A correction voltage will always be present and will be of the correct polarity.

The foregoing description of the tuning servo is based on the resonant resistance of the cavity being greater than the characteristic impedance Z_0 of the line. If the resonant resistance is less than Z_0, the minimum position changes from the S-shaped curve of Fig. 12(A) to a discriminator type of characteristic as shown in Fig. 12(B).

Under this condition, the probes are set to straddle a maximum at resonance. Output polarity for a given direction of standing-wave shift is opposite to that in the previous case. However, the slope of the standing-wave shift with load reactance is also reversed, so the net effect is still in the proper direction for correction.

The system just described has been built and tried out with various dielectric loads, using the cavity previously described. The equipment is shown in Fig. 13. The servo amplifier used is a Brown Electronik unit, with a matching Brown servo motor for tuning the cavity. The phase detector consists of two adjustable Hewlett Packard probes, with a Microlab *XM-1000* crystal detector mounted on each probe. The probes are mounted on the slotted line assembly with a quarter-wave spacer between them, so that the positions of both probes can be adjusted simultaneously. Matched crystals of opposite polarity were used, and the sensitivities of the probes were equalized by adjusting probe depth.

Operation of the system was highly satisfactory when heating materials with loss factors of 0.03 or greater. Probe position was found to be somewhat critical, requiring that the probes be set within a range of 2.0 millimeters at 915 mc. Once set, however, the probes required no further adjustment. The cavity would come into resonance as soon as power was applied, from any tuning condition whatever. Tuning was found to track the change in the sample very closely. The match of the cavity to the line was reasonably good, with a maximum VSWR of 2.0 for most samples. The cavity coupling still had to be matched to the sample impedance when changing sample material or sample size, but did not need to be changed during the heating of any particular type of sample.

Operation of the system with low-loss materials (loss factor below 0.03) was more critical and less satisfactory. The system tended to hunt when the probe position straddled the minimum at resonance, and the probes had to be moved slightly along the slotted line before the system would stabilize. This adjustment destroyed the ability of the system to seek correct tuning from all initial tuning conditions, but would hold nearly correct tuning once it was achieved. The trouble appeared to be caused by magnetron frequency jumping rather than any servo system malfunction, since the same unstable characteristics were present when the cavity was manually tuned. This effect has not yet been eliminated and further work is indicated.

Conclusions

An investigation of the characteristics of rapid heating of various dielectric materials at 915 mc shows that at this frequency materials can be heated which cannot be heated at conventional preheating frequencies below 70 mc. A detailed investigation of phenolic preforms shows that it is possible to design a heating cavity which will provide rapid heating with reasonable uniformity. The problem of impedance matching requires special consideration, since the heating cavity is not the primary resonant circuit as in the case of a conventional preheater operating in the very-high-frequency band. A servo-tuned cavity has been devised which serves to keep the cavity reasonably well matched to the magnetron output during the heating cycle. This technique provides a satisfactory solution to the matching problem for materials having a loss factor greater than 0.03. For lower loss materials, an initial tuning adjustment was necessary.

References

1. Dielectric Nomenclature (Letter to Editor), H. H. Race. *Electrical Engineering*, vol. 61, no. 2, 1942, p. 105.

Some Unusual Designs of Electric Resistance Heating

LEE P. HYNES
FELLOW AIEE

LARGE installations of electric heating depend for their success on a correct analysis of the many technical problems involved. Too often these are not given the proper study and unsatisfactory operations result from inadequate engineering design.

Some unusual installations of resistance heating will be described and the technical problems, including heat transfer, will be analyzed. As a refresher, some of the basic inherent advantages of resistance heating and the laws of electric heat transfer, will be briefly discussed.

Resistance Heaters

Resistance heaters have a unity power factor and an efficiency of 100% in the conversion of electric energy into heat. There are no inductive losses and one kilowatt-hour can be counted on to produce 3,415 Btu. Some loss is unavoidable during the transfer of the heat to the work material, but this can be reduced to a minimum by careful design and good thermal insulation.

Nickel-chromium resistors have a low temperature coefficient and will never draw more than about 7% added inrush of current, when starting up cold. Circuits do not need extra large conductors and can be protected by normal fuse ratings, without overload time lag features.

Resistance heaters can be designed to meet all code requirements for use in hazardous atmospheres and for all standard commercial power lines, without special transformers.

Heat Transfer

Too little consideration is usually given to the transfer of heat from the resistors which produce it to the work material. This heat transfer is the key to a successful installation of electric heat, especially when the operating requirements are new and difficult. Heat is primarily transferred from a resistor by conduction and by radiation but can be further distributed by convection.

Conduction passes heat from one part of a material to another, or from one material to another which is in physical contact with it. This transfer does not involve any displacement of the body particles. Different materials vary greatly in their ability to conduct heat.

Convection distributes heat energy within the body of a fluid by thermal currents or by forced circulations that displace body particles and cause thermal mixing. The flow of fluid over a heating surface wipes away film and improves heat transfer. This reduces the ΔT and results in a lower temperature of the heating surface, but with an electric heater this does not change the rate of heat input.

Radiation disperses heat energy in all directions by electromagnetic waves of high velocity. When two bodies of differing temperatures, for example a heater and the work material, are located within an enclosure, there is a continual interchange of radiated energy between them. The hotter body radiates more energy than it absorbs and the cooler body absorbs more energy than it radiates.

Radiant heat rays have differing wave lengths, according to the temperatures of their sources. This gives them differing powers of penetration into various materials, depending upon the absorptivity of each material.

Some materials are transparent to short-wave radiations from high-temperature sources, such as heat lamps, which have filaments running up to 4,000 F (degrees Fahrenheit). Ordinary glass is a good example of a material that cannot be effectually heated by such short-wave radiations, as the heat waves pass straight through it and are not absorbed. Such a material can be satisfactorily heated by longer waves from a source in the 700 F range.

The efficient application of radiant heat requires a full understanding of the absorptive qualities of each material and the wave length that will give the best results. Most common gasses such as air, nitrogen, hydrogen, and oxygen will not absorb radiant heat, but vapors are good absorbers of radiation.

The problems of heat transfer must be carefully considered in designing any large industrial heating project. Too many engineers seem to think only in terms of steam heat transfer, and fail to realize that electric heat operates on radically different basic principles. Steam has a fixed temperature, for a given pressure. When the work is cold, the condensation rate is high, but this tapers off as the work temperature rises, and the rate of heat input falls because the ΔT is reduced.

Electric heat operates in a different manner because the rate of heat input (wattage) remains constant, for a given voltage, regardless of the work temperature. As the latter rises, the heater temperature also rises and there is no reduction of the ΔT, because all of the heat energy produced by the electric heater must be disposed of, regardless of the temperature required to do so.

This basic characteristic of temperature change in an electric heater, has an important bearing on every electric heating application, because materials vary so much in their inherent ability to absorb heat, without damage, from contact with hot surfaces.

This makes it necessary to understand and provide for a proper heat density, in watts per square inch of heating surface, for each material. Molasses and sugar are good examples of materials extremely sensitive to heat; they readily caramelize and burn on hot surfaces at comparatively low temperatures.

The heat capacity of each material, its conductivity, and the mass being heated, also play a large part in the proper design of an electric heated process and its automatic control. Large masses change temperature slowly, and tanks of oil or water, for example, seldom are difficult to control. Where materials are passed rapidly through heating zones, especially at variable speeds, the control problems become more complicated.

The many special characteristics of electric heat offer the alert designer many opportunities for producing new designs that will obtain surprising results and that are not possible with other types of heat sources. A few examples of unique installations of resistance heaters will be described, and the heat transfer which is involved in each will be analyzed. These will show how important heat transfer is for securing the desired operating results. A factor of great importance is the correct utilization of radiant heat waves, even

Paper 59-77, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 30, 1958; made available for printing December 9, 1958.

LEE P. HYNES is a consulting engineer, Haddonfield, N. J.

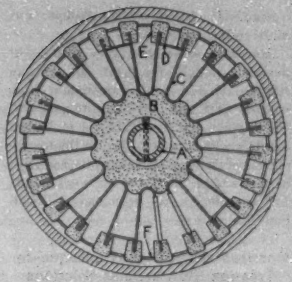


Fig. 1. Cross section of Radi-Fin electric heater for gases

when the material being heated does not directly absorb radiation.

Heating Hydrogen and Other Gasses

The manufacture of ammonia presented a problem requiring the heating of 2,400 pounds per hour, of hydrogen, at a pressure of 3,000 pounds per square inch. The required temperature was 1,400 F and the space limitations were very severe. Because of the high pressure, it was essential to keep the pressure vessel diameter down to 6 inches inside diameter. The heated length could be only 14 feet. The required heat input was 400 kw at 440 volts, single phase. This job posed problems for which little precedent was available as a guide.

The internal wall of the cylinder had a surface area of only 22 sq ft (square feet) and to use that wall as the sole heating surface, with an external electric heater, would require a heat density on this surface of 125 watts per square inch. Such a high rating would require a very high ΔT, in order to force so much heat energy through the wall of the cylinder and into the gas it contained. It was at once obvious that a much greater area of heater surface would be required and that it must be located within the gas stream, where all the heat could be transferred to the gas by conduction, as the gas wiped over the heating surfaces. As previously mentioned, hydrogen does not absorb radiant heat.

The great importance of heat density has been mentioned but in this case the hydrogen would not be harmed by contact with high-temperature surfaces. The problem was to keep down the temperature to a safe value for the wall of the cylinder and for the heating resistors. The life of a metallic resistor is long, up to a certain operating temperature. Then it decreases fast and at an accelerating rate, as the temperature is further increased. As the temperature of an electric resistor rises until it can deliver all of its rated heat output, the abnormal rating of 125 watts per square inch, would have greatly overheated the resistor and reduced its life to a serious degree. Surface density is a factor of surface temperature.

Unfortunately, the suitable heat density rating for different materials, under different operating conditions, has never been sufficiently explored scientifically. Even the manufacturers of resistor materials did not know what would be a safe rating under the severe operating conditions outlined for this job. Some attempts had been made to heat hydrogen with round rods as resistors, but the resistors had lasted but a very short time. Computations clearly showed why they had failed as they had a heat density of over 150 watts per square inch, which is a fantastic figure. This was the only lesson available from practical experience, so a new start had to be made.

It was decided to approach the problem in reverse order, by first determining the maximum area of heating surface that could be designed into the available space, and then see what amount of heat density would result. The obvious location for the heating surface was inside the pressure cylinder, where the gas would wipe directly over it. This was feasible as hydrogen is not a conductor of electricity.

Much thought was given to finding a geometrical shape of heating surface that would have the maximum area and still offer the least resistance to the flow of gas through the vessel. The unique design shown in Figs. 1, 2, and 3, was finally developed. This had over 60% of free area for gas flow and the radial fin construction had a total prime heating surface area of 85 sq ft. This was nearly four times the internal wall surface of the vessel and it resulted in a heat density of only 32 watts per square inch.

It was possible to reduce this still further by utilizing radiant heat. Although hydrogen does not absorb heat radiation waves, the waves are radiated through the gas to the internal wall of the vessel, which they heat. As the gas wiped this wall in passing longitudinally through the cylinder, the wall became a very good secondary heating surface, with an area of 22 sq ft. This brought the actual average heat density down somewhat below the 32 watts, as given in the foregoing for the heating resistor. Several years of continuous service has proved that the rating was a safe and conservative one.

A warning should be given here that there is no standard safe heat density, for even a single material. Heat density depends on several factors, such as the specific heat, the density and flow rate of the gas, and the final temperature to which it is to be heated. For heat absorption, hydrogen has the best combination of favorable factors.

Similar designs have been successfully used for heating air and other gases, but in each case the heat density was computed for the gas and the operating condi-

Fig. 2 (left). Complete assembly of Radi-Fin heater for gases

Fig. 3 (right). Radi-Fin heater being inserted into a high-pressure vessel

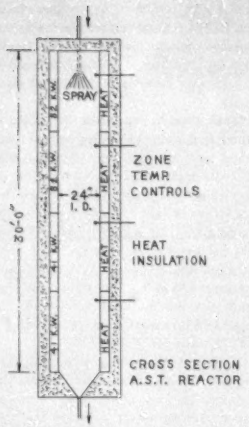


Fig. 4. Cross section of an AST process vessel for treating sewage with electric heat

tions. Although good results have been obtained by careful estimating, and conservative ratings, it is to be hoped that the subject of over-all heat transfer rates for electric heaters, in gases, liquids, and other materials, will eventually be more thoroughly studied scientifically. This would provide better design data for engineers.

Returning to the design details, Fig. 1 is a cross-sectional drawing. The heating fins were installed around a tubular backbone member, *A*. This was not alive electrically; it was supported rigidly by brackets at the terminal end. The other end was guided but was left free to expand and contract freely without putting any strain on the fixed terminal connections. At intervals along the tube *A* are loosely mounted gear-shaped refractory insulators *B*. Each insulator has 12 curved recesses in its perimeter. In these recesses were seated 12 U-shaped resistors *C*, formed of nickel chromium furnace ribbon 3 inches wide.

The 12 U-shaped resistors were electrically connected in series by jumpers *D* for single-phase. For 3-phase, each phase would consist of four resistors in series to an internal Y bus. For single-phase, two conductors are brought out through one end of the vessel, and for 3-phase there are three such conductors. Special high-temperature, high-pressure terminals were developed for this hard service. With the terminals located at one end of the vessel, it is possible to introduce or to remove the complete heating element, by removing one vessel head. This does not disturb the general setup, nor interfere with the thermal insulation which surrounds the vessel walls.

Fig. 1 also shows a series of small refractory insulators *E* which fit the outer edges of the fins of the U-shaped resistors *C*, and form a circle. They are held in place by an external band *F* of nickel chromium that lies in grooves of the insulators *E*. These circles of insulators are located at intervals along the heating element and serve to keep the outer edges of the U-shaped fins, properly spaced from each other and from the wall of the vessel. Fig. 2 shows a complete heating unit and Fig. 3 shows the unit being inserted into the 6-inch pressure tube enclosure. It also shows the 3-phase terminal rod connections.

The entire assembly is very rugged, yet flexible enough to meet all normal stresses of high temperature and high velocity of dense, high-pressure gas flow. There were many mechanical and electrical problems to be solved in the manufacture of this new type of heater unit, but by care and patience they were overcome. The new design has been named "Radi Fin."

Atomized Suspension Technique for Treating Slurries, and Solutions

The heating problem under discussion arose from a research program which had been established to find some way to recover sulphur from paper pulping liquor, which for disposal had heretofore been poured into streams and thus wasted. The lost sulphur was costly and the waters were badly contaminated. Preliminary tests by the research staff indicated that no standard commercial electric heaters would endure under the severe operating conditions demanded, and a design engineer was retained to study the problems involved.

For the new process with which the scientists planned to experiment, an internal temperature of about 1,500 F was specified. To prove their theory of preliminary drying, with a following radiation reaction to destroy the solid particles by combustion, the scientists were willing to accept an electric heating system which might be expected to endure even for a few hours only.

No heating equipment could be installed inside the vessel and the only possible heating surface was the actual wall of the vessel itself. Electric heaters would

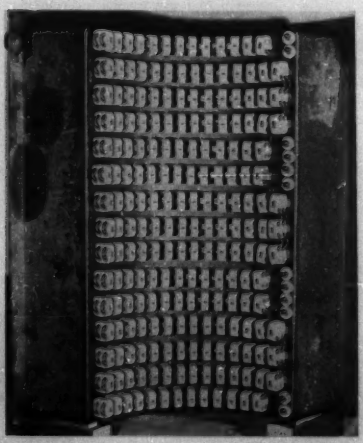


Fig. 5 (left). Radiant electric heater section for an AST vessel

Fig. 6 (right). Field assembly of an AST process vessel

have to be mounted outside this wall and radiate heat to it. A design was developed with four heat zones, in vertical positions, with separate temperature control for each zone. With this system it was possible to recover over 70% of the sulphur from pulping liquor and the heaters were so durable that they survived for many subsequent experiments on other compounds.

The process was found to have wide application and it has been patented in many countries. Many firms are licensed to use it for a number of important industrial processes in the treatment of slurries and other solutions. The "Atomized Suspension Technique," (AST), as the process has been named, was worked out by Prof. W. H. Gauvin of McGill University and is owned by the Pulp and Paper Research Institute of Canada. The original vessel was only 10 inches in diameter and approximately 8 feet high. Following the success of this small unit, several larger ones have been constructed for further experiments.

An unexpected discovery was that sanitary sludge from raw sewage could be successfully treated and destroyed by passing it through one of these AST process units. The first large units for this service have been designed and installed in a Canadian city and if the results in actual daily use equal the experimental tests, a new field for sewage treatment by electric heat may be opened up.

Fig. 4 shows a vertical cross section of one of these sewage-treatment vessels, now being installed. Its inside diameter is 24 inches and it is 30 ft high. The total electric load is 246 kw with four vertical heat zones. The temperature of each zone is independently controlled. The two upper zones are rated at 82 kw each and the two lower at 41 kw each. Each zones has a complete 3-phase 575-volt electric heating unit. Fig. 5 shows one of the three single-phase heating units which comprise each heat zone.

The three heating units for each zone are arranged with one on each phase. These heat sections are removable from the main vessel without disturbance to any other zone. Each section has its own structural frame which carries a series of refractory insulators arranged horizontally in parallel arcs. They support heavy helical resistor coils which are free to expand horizontally, but which cannot sag under high temperature.

Fig. 6 shows the field assembly of one of these sewage-treatment units. The internal work cylinder is shown with the four separate heat zones. Each zone is also divided into three compartments by vertical partitions, one for each of the single-phase heater elements. The inner chamber is free to expand and contract independently of the external structural supports. Heating units as shown in Fig. 5 have been installed in the two lower heat zones. The heaters for the two top zones may be seen on blocks in the foreground, awaiting installation.

The heat sections are mounted to vertical supports in such a way that they may expand freely. Metallic reflector shields surround the heaters and an external layer of suitable thermal insulation conserves the heat energy.

The method of operation is to introduce the fluid material into the top of the vessel as shown in Fig. 4, in the form of a fine spray. As it travels downward it is subjected to a drying process that evaporates most of the moisture. Solid particles float toward the bottom and are further acted upon by radiation, which destroys organic matter and converts it into gases and some fine ash.

At the bottom of the vessel all the treated products are vented through separators and other apparatus. Some of the waste heat can be recovered, and the gases and ash disposed of in a safe and sanitary way.

Engineering Considerations

The brief description of these two processes should serve to point out the opportunities for special designs of resistance heaters, in the solution of industrial problems. There is also need for more comprehensive data on basic electric heating information. Engineers should recognize the importance of bold and imaginative approaches to new and difficult heating problems, in which electric heaters may serve in a unique way.

Anode Breaker Testing with a High-Capacity Rectifier System

L. J. HARRIS
NONMEMBER AIEE

T. J. SCULLY
ASSOCIATE MEMBER AIEE

V. N. STEWART
MEMBER AIEE

Synopsis: An extensive testing program for anode breakers using a high-capacity power-rectifier system of the Aluminum Company of America was an unusual opportunity for an appraisal of the combined performance of the system equipment. The information gave an evaluation of the protection to the equipment and has provided a better understanding of the influence of the system components on operating service conditions. This was done with special controls and comprehensive instrumentation which achieved a uniform testing procedure including the rectifier-tank characteristics, phase control of rectifier-tank firing, and interphase transformer saturation control. Through the repetitive consistency of the test arc-back currents, a reliable performance was given by the high-speed current-limiting breaker over the range of available power. In addition, these tests with the large aluminum potline rectifier system gave a real measure of the combined performance of the equipment up to the maximum service requirements rather than by individual device tests, the data for comparison and evaluation of other test facilities, and a guide to future applications.

POWER CONVERSION systems for electrochemical processes including aluminum, chlorine, and sodium, railway, mines, and general industrial applications are important loads on the total electric power generation. Not only are these applications using larger concentrated blocks of power but are striving to achieve higher over-all efficiency in the process operations in the electrochemical industry. Likewise, there is a growing use of d-c power for the greater versatility and control which can be attained by d-c devices. The use of d-c motors has increased percentagewise several times that of the a-c motors in sizes suitable to machine-tool applications during the past 10 years.

Paper 58-224, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., November 2-7, 1958. Manuscript submitted November 6, 1957; made available for printing September 12, 1958.

L. J. HARRIS is with the Aluminum Company of America, Rockdale, Tex.; T. J. SCULLY and V. N. STEWART are with the General Electric Company, Philadelphia, Pa.

The authors acknowledge the assistance of A. Schmidt, Jr., and W. G. Abbott of the General Electric Company, and of the engineers of the Aluminum Company of America (Alcoa), Rockdale Works, in carrying out the successful test program.

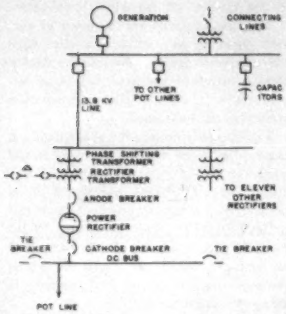


Fig. 1. One-line diagram of Aluminum Company of America mercury-arc-rectifier system

Such systems must be reliable, rugged, dependable, and efficient to maintain an uninterrupted flow of d-c power. Field testing of components such as the anode breaker will assure the reliable performance of the equipment and the adequacy of the application.

The anode breaker plays an important part in the system where power disturbances are caused by arc backs in the power-rectifier tanks. Arc backs are a failure of the rectifier action and result in the flow of current from the cathode to the anode within the rectifier tank. This is equivalent to a short circuit on the secondary terminals of the rectifier transformer, and the magnitude of the current contributed through the rectifier transformer, and the magnitude of the current contributed through the rectifier is determined by the impedance of the a-c system. In addition, on these high-capacity systems, there will be a d-c bus feedback of an exponentially increasing current due to the associated rectifiers, generators, and the loads. The magnitude of this current and the ability of this breaker to open the fault with minimum loss of forward power are important factors in achieving a dependable system.

This tank which is conducting current in the reverse direction will have an initial peak current after a time of 0.009 second.[1,2] For high-speed breakers, this first-crest arc-back current is the point of interest, as it represents the maximum duty during interruption. Further, the high-speed operation of the anode breaker gives current-limiting action to minimize this peak current and reduces the stresses from the short-circuit duty on the associated equipment.

Power Rectifier System

The system used in this test program is represented in the 1-line diagram, Fig. 1, which shows the normal components of generation, transmission, and transformation feeding an aluminum potline through 12 double-Y-connected mercury-arc rectifiers. The major source of power is 300,000 kva of generation and it is supplemented by three tie lines connecting with the local power utilities.

The normal operating duty and arc-back conditions handled by the breakers and other components are related to the capacity of 12 double-Y-connected mercury-arc rectifiers which are associated with one potline. However, the system arrangement provides bus ties with adjacent potlines for emergency operation. The tie breakers could be closed and during this abnormal connection greater arc-back currents would be obtained.

Fig. 3. Detail of test drawout mechanism for anode breaker

Test Arrangements and Facilities

The rectifier unit selected for this program was located at the end of a line-up of 12 identical units with an equivalent line of high-capacity rectifiers extending in the opposite direction from the test location. Tie breakers permitted the combined operation of these and other potline busses.

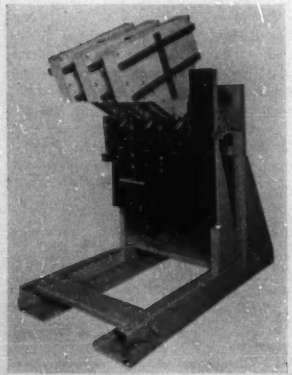


Fig. 2. Three-pole type AG-3 drawout anode breaker and frame used for arcback tests

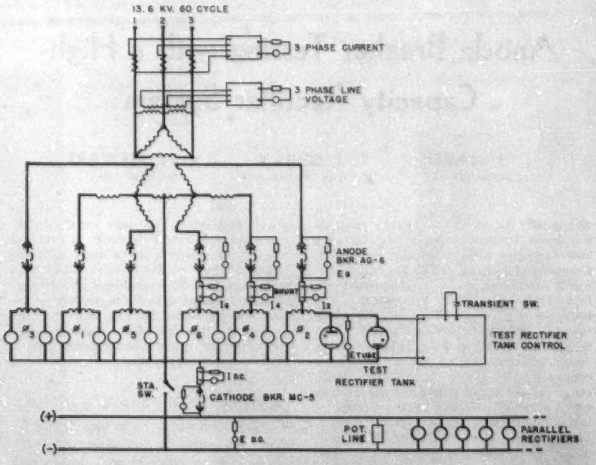


Fig. 4. Test connections on mercury-arc power rectifier

Fig. 5. View of instrumentation and control in test area

Two test anode breakers, Fig. 2, for 6-pole operation, were inserted in the system in place of the normal station equipment and controls were provided for single-pole operation. This type of operation kept the rectifier supplying the power load and the phase or phases opened by arc backs could immediately be restored to service. The drawout features used on these anode breakers, Fig. 3, made them easily removable from the system during the testing program for rapid inspection and such changes as were necessary to vary the time of tripping during system measurements.

A pumpless rectifier tank, 16 inches in diameter, was connected in parallel with a phase-2 rectifier tank as shown in Fig. 4. This test-rectifier tank with throw-over facilities could be operated to carry load or to give arc-back current. By this means normal arc-back characteristics of the tank were present in the circuit for the equipment tests.

A full-phasing control unit operated in conjunction with a selsyn and provided the necessary synchronous timing for the firing pulse of the ignitor. The single-short feature is obtained in the control by limiting the ignitor pulse to the energy stored in a capacitor. Blocking of further firing pulses is obtained by the initiating relay which prevents renewal of the charge on the capacitor until a reset sequence had been completed. The phasing-control circuit in turn received initiation from the magnetic oscillographs. These oscillographs and the other test equipment are shown in Fig. 5. Experience confirmed the highly accurate and consistent timing of the rectifier-tank firing during the entire testing program. The majority of the tests were made to give commutation arc backs which subjected the anode breaker to the maximum service condition.

The measurements of recovery voltages were recorded with three types of instrumentation including the cathode ray, surge-crest voltmeters, and the magnetic oscillograph. This permitted full observation for all possible peak voltages that existed in the test program. An agreement was found between all methods used.

Interphase Transformer Saturation Control

During the planning stages for the test, it was recognized that a wide variation could occur in the arc-back current due to the recent past magnetic history of the interphase transformer. It could be reasoned that the volt-seconds of the interphase transformers have varied in previous test programs so that an equal comparison between even sequential tests was difficult. A rectifier subjected to the varying conditions of the test program could readily produce an unbalance through such conditions as a rectifier-tank misfire. Therefore, a means was provided to saturate the core of the interphase transformer in a predetermined direction by the deliberate removal of the current in one phase which simulated misfire operation. The resulting unbalanced magnetization of the interphase transformer was sufficient to carry the core iron into a maximum saturated condition. Immediately before the initiation of the arc back, the circuit was allowed to operate normally as shown by the return of the load-current loop of phase 6 in Fig. 6(A). In addition to the direction saturation of the interphase transformer, this allowed phase 6 to contribute to the arc-back current in a normal manner.

The location of the ignitor-blocking means on a rectifier tank in the same Y as the arc back, or in the other Y, determined whether the interphase transformer core was conditioned to give maximum or minimum arc-back current. To demonstrate the opposite effect of the interphase saturation, the ignitor-blocking means was shifted to phase 5. This

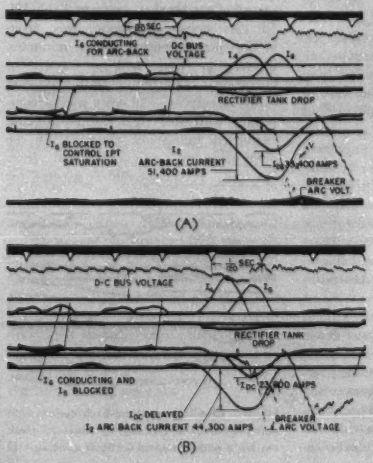


Fig. 6 (left). Oscillogram of arc-back current on normal system with interphase transformer saturation controlled to give delays in d-c contribution

A—Minimum delay: circuit condition B

B—Maximum delay

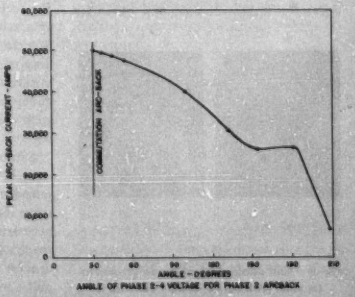


Fig. 7 (right). Variation of peak arc-back current with initiating angle

Table I. Arc-Back Tests with Interphase Transformer Saturation Control

Test No.	Oscillogram, Fig. No.	Circuit Condition	Breaker Operating Speed	Saturation Direction	Rate-of-Rise, Amperes Per Second	Peak Arc-Back Current, Amperes	Time to Peak, Milliseconds	Peak Current From D-C Bus, Ampere
4	5	B*	High Speed	For maximum rate-of-rise of arc-back current	8.5×10^8	51,400	8.4	33,400
2	6	B	High Speed	Reversed	7×10^8	44,300	7.9	23,600
130	7	B	Delayed	For maximum rate-of-rise of arc-back current	8.0×10^8	56,300	9.2	33,400
125	8	B	Delayed	Reversed	6.75×10^8	47,300	9.0	22,300
159	12	D	Delayed	For maximum rate-of-rise of arc-back current	8.9×10^8	63,500	9.0	36,100

* Condition B is normal.

Table II. Circuit Conditions for Control of Available Power

Circuit Conditions	Potlines No.	Special Air-Core* Reactor	Rectifier Transformer Reactor	Measured Available Peak Arc Back, Amperes
A	3	In	In	
B	3	Out	In	55,000
C	2 & 3	Out	In	
D	2 & 3	Out	Out	63,500

* Used in place of anode-dividing reactor.

created a condition giving maximum flux change to oppose d-c fault current through the interphase, and the reverse bus current was delayed, as shown in Fig. 6(B), which resulted in a lower peak magnitude.

Table I gives measured values of the effect of the magnetic conditions of the interphase transformer associated with rectifier-tank misfire performance. With normal system condition and using a delayed breaker, the saturation control varied the peak arc-back currents 19% and the rate-of-rise 17%. The difference in the peak current reached by the reverse feed from the d-c bus was 11,000 amperes and represents a 50% variation. These variations without control would cause considerable error in the appraisal of the breaker's current-limiting ability. Arc-back tests were repeated many times with the interphase magnetization controlled to produce maximum arc-back current for each power level. This procedure with phasing control did remove variations and gave reproducible results in all tests.

Arc Backs at Various Initiating Angles

While the commutation arc back produces the largest current and has been recognized as the most severe duty for breakers, arc backs do occur at other conditions. However, to assure a complete investigation, tests were made with a complete range of initiating angles. The selsyn phase control on the firing of the test rectifier tank provided the means for varying the arc back. The peak values of arc back current are shown in Fig. 7 with reference to the phase-2-4 voltage. The delay in arc-back initiation was increased until the arc back current was of such a small value that it tripped the anode breaker without a visible flash and very little noise. Hence it is possible to have a "silent arc back" and to open the anode-breaker pole. A picture of the interruption of a normal commutation arc back is shown in Fig. 8.

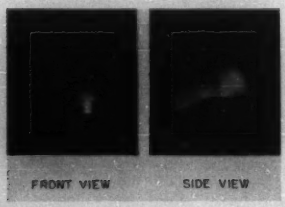


Fig. 8. Still photograph of anode breaker interrupter with normal system conditions

Measurement and Control of the Available Arc-Back Current

The available power could be set at four levels for the test program. This gave a range which included a value less than normal and two values greater than normal operating duty. The first step above normal level of operation was obtained by connecting two potline busses in parallel. The second level above normal was obtained by a short circuit around the rectifier transformer reactor in addition to connecting two potlines in parallel. For the level less than normal, a special air-core reactor was used in place of the normal anode-dividing reactor. These circuit conditions are shown in Table II with code for easy use with other references.

Establishment of the available first crest in the arc-back current was determined by test for the normal and maximum conditions. For this test the magnetic trip circuit of the anode breaker was modified to delay the opening of the anode breaker pole. The arc-back current under maximum system conditions reached a first-crest value of 63,500 amperes which is a lower current than determined by many of the past methods of calculation. This value using a rigorous solution has been checked by a method given in two companion papers.[1,2] The oscillogram for the maximum test level is shown in Fig. 9.

Anode-Breaker Field-Test Performance

The entire test program resulted in more than 200 controlled arc backs on the potline. This program was associated with a single rectifier unit and the anode breakers under test were located directly in the circuit to provide normal power to the load. Except for a few special tests, interrupting duty was limited to a single phase of the system. In comparison with normal system operations such a series of tests represents 50 years of arc-back duty and gives a good record for design of the whole power-conversion system.

Representative of the anode-breaker performance in all these tests is the field record of one breaker shown in Table III. In this series, a measure was given of the current-limiting ability, minimum interrupting time, and consistency of operation over the minimum-to-maximum range of current. All tests were accomplished without system disturbance and permitted immediate reclosing of the open phase for continuous output from the unit. A typical film is selected for

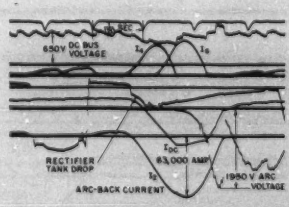


Fig. 9. Oscillogram of arc-back current with delayed breaker tripping to measure peak available under maximum circuit condition D

Table III. Field Test Performance for a High-Speed Anode Breaker

Test No.	Circuit Condition	Measured First-Crest Available Current, Amperes	Peak Arc Backs, Amperes	Total Clearing Time, Milliseconds	Peak Arc Voltage	Rate-of-Rise, Amperes Per Second
1*	B	55,000	50,000	13.9	1,480	8,270,000
2*	B		44,300	13.0	1,830	6,780,000
3*	B		44,300	13.1	1,630	7,000,000
4*	B		51,400	13.2	2,100	8,480,000
5*	B		51,400	12.8	2,030	8,500,000
6*	A		45,700	12.0	1,780	6,890,000
7*	A		45,000	12.2	2,000	7,220,000
8	A		44,300	12.2	2,100	6,480,000
9	A		44,300	12.4	1,833	6,180,000
10	A		44,300	11.9	1,933	7,230,000
11	B	55,000	51,400	11.9	2,200	7,500,000
12	B		51,400	11.8	2,200	8,240,000
13	B		51,400	12.3	2,100	8,330,000
14	B		51,400	12.1	2,200	8,460,000
15	B		50,400	12.2	2,100	8,140,000
16	C†		50,700	12.1	2,270	8,190,000
17	C†		51,500	12.2	2,200	8,440,000
18	C†		51,500	12.4	2,270	8,480,000
19	C†		51,500	11.9	2,270	8,260,000
20	C†		51,500	11.9	2,130	8,320,000
21	C†		58,600	12.2	2,270	9,930,000
22	C†		55,800	11.5	2,230	9,470,000
23	C†		57,300	12.2	2,230	9,490,000
42	C		52,900	12.9	2,465	8,720,000
43	C		52,900	11.8	2,430	8,640,000
44	C		52,900	11.4	2,465	9,000,000
45	C		52,900	11.5	2,100	9,240,000
46	C		52,200	11.5	2,300	8,960,000
47	D	63,500	57,250	11.5	2,065	10,100,000
48	D		56,500	11.6	2,300	9,230,000
49	D		57,250	11.5	2,165	9,790,000
50	D		58,600	11.6	2,230	9,810,000
51	D		55,700	11.6	2,400	9,370,000

* Preliminary tests to explore saturation control.

† Parallel potlines 3 and 4 instead of 2 and 3.

Table IV. First-Crest Arc-Back Currents with Various System Conditions

Test No.	No. of Generators	No. of Connecting Lines	Potline No.	No. of Capacitors (7,500 Reactive Kva)	Transformer Reactor	Peak Arc-Back Current	Time to Peak Current, Milliseconds
147	3	2*	3	4	in	55,000	8.9
152	3	2*	3	0	out	61,800	8.7
153	3	0	3	0	out	61,500	8.6
155	3	0	3	7	out	63,000	9.2
157	3	2*	3	7	out	61,500	9.2
159	3	2*	2 & 3	7	out	63,000	9.6
161	3	2*	2 & 3	0	out	63,000	9.1

* Power out.

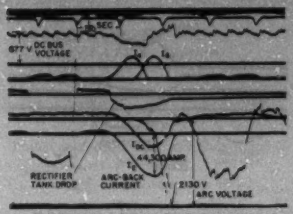


Fig. 10. Oscillogram of arc-back current with special anode reactor for less than normal available current (circuit condition A)

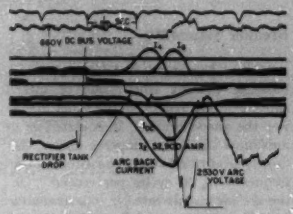


Fig. 11. Oscillogram of arc-back current with parallel d-c busses for greater-than-normal available current (circuit condition C)

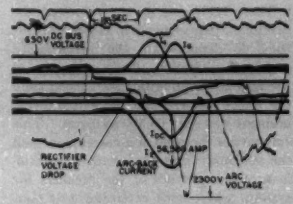


Fig. 12. Oscillogram of arc-back current with parallel d-c busses and rectifier transformer reactor omitted for maximum available current (circuit condition D)

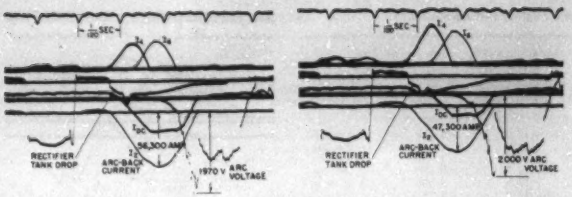
each level of available current, Figs. 6, 10, 11, and 12, and includes the record of all phase currents in the Y under test, the d-c feedback, arc voltage of breaker, voltage on the rectifier tube under arc back, and d-c bus voltage. These oscillograms show the controlled misfire of phase 6 to assure maximum arc-back currents for each level tested. The consistency of anode-breaker performance over all levels is represented by a 3% variation in let-through peak arc-back current and a 14% variation in total clearing time. The absence of any restrikes at any level of operations makes possible the consistent values of total clearing time.

System Measurements

The test program had a dual purpose, and both objectives are significant in establishing adequate knowledge and performance of mercury-arc-rectifier systems. In addition to anode-breaker performance, system measurements were made to determine actual service condition and the influence of each component.

To obtain the desired information, the anode breaker was modified to delay the opening of the contacts until the first crest of arc-back current was reached. This delay was used for combined system measurements as well as those special tests conducted to establish the independent contribution of the d-c bus, single-Y and double-Y connections. Oscillograms, Figs. 13 and 9, are given for the normal and maximum levels with corresponding crest arc-back currents of 55,000 and 63,500 amperes using circuit conditions shown in Table II. This is appreciably less than a calculated value of over 75,000 amperes for these systems using previous methods and gives a better basis for establishment of the system requirements.

In addition to the control of the immediate circuit with reactors, the system conditions were varied and the first-crest arc-back currents measured. The results indicated in Table IV are based on repeated tests in all cases with identical results. Further, it is evident that the influence of adjacent potlines through bus ties, connecting transmission lines, and capacities banks are small or negligible compared with the impedance changes in the individual rectifier-tank circuits and transformers.



A—Minimum delay in d-c contribution

B—Maximum delay in d-c contribution

Fig. 13. Oscillogram of arc-back current with delayed breaker tripping to measure peak available current on normal system with inter-phase transformer saturation controlled to give delays

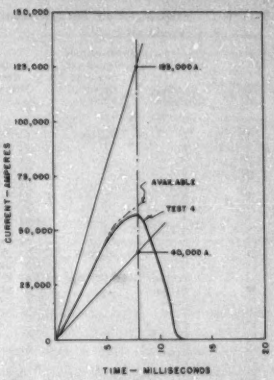


Fig. 14. Interrupting performance of breaker compared with National Electrical Manufacturers standard for high-speed breaker applications

Breaker Standards

Present-day standards[3] for high-speed breakers present two forms of interrupting rating based on the application. For the anode breaker, it is stated as a maximum of 75,000 amperes during the first cycle at rated voltage based on supply frequency of 60 cycles per second. On the d-c air circuit breaker, the available current at 0.008 second should be between 40,000 and 125,000 amperes at rated voltage and the breaker speed sufficiently fast to limit the crest current by a time of 0.010 second.

These field tests do not represent the maximum requirements in either case but they do demonstrate the performance within these limits. Early estimates of the system available indicated higher than the 75,000 amperes but the actual measurement was only 63,500 amperes. This emphasizes the need of a correlation with the simulated tests to demonstrate satisfactory performance of the breakers for systems where such extensive tests cannot be conducted.

In the field tests against high-speed d-c air circuit breaker requirements, Fig. 14, the inherently fast anode breaker readily demonstrates its ability for this function even though the rate-of-rise is varying with the a-c system. The high speed makes it a current-limiting device and limits the peak well in advance of the first crest of the system and within the 0.010-second time of the standard.

Conclusions

The extensive testing program conducted on the high-capacity mercury-arc rectifier system of the Aluminum Company of America provided considerable information about the system and anode breaker performance. These conclusions were reached:

1. Testing means, which gave conditions comparable to the service conditions, could control and repeat the arc-back current through the anode breaker with extremely good consistency as indicated by very small variations of 6% in peak current observed in the anode breaker tests. Uncontrolled, the nonlinear impedance of the interphase transformer could cause variations of 50% in the d-c contribution, and this alone amounted to a variation of 19% in the total arc-back current.

2. Consistent high-speed current-limiting action of the breaker was demonstrated. This program included a range of duty and as many as 60 operations on a single-pole unit were performed. The current through the breaker was only 90% of the maximum available first-crest current.

3. Measurement of the available first-crest arc-back current levels were made with delayed breakers and the results were in full agreement with new calculations. The 63,500 amperes measured for the system is under the 75,000 amperes maximum for the anode-breaker standard.

4. The contribution of the d-c bus to the first-crest arc-back current is approximately 65% for normal system operation with a transformer reactor, and 45% of the total when the reactor is omitted.

5. Variations due to the random magnetic conditions in the interphase transformer were eliminated and precise control gave a maximum arc back during all performance tests.

6. Factors influencing the magnitude of arc-back current have been measured. The greatest changes are associated with the rectifier tank circuit and transformer impedance. This information can be used to advantage in planning future designs and applications.

References

1. Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers, H. P. Fullerton, J. Teno. *AIEE Transactions*, vol. 77, pt. II, Nov. 1958, pp. 456–64.

2. Comparison of Calculated and Measured Arc-Back Currents in Large Power Rectifier System, J. Teno, C. H. Titus, R. N. Wagner. *Ibid.*, 1958 (Jan. 1959 section), pp. 585–89.

3. Large Air Circuit Breakers. *NEMA Standard SG3-1955*, National Electrical Manufacturers Association, New York, N. Y., 1955.

Relay-Type Feedback Control Systems with Dead Time and Sampling

KEISUKE IZAWA
NONMEMBER AIEE

LYNN E. WEAVER
ASSOCIATE MEMBER AIEE

THIS investigation is a time domain study of relay-type feedback control systems with dead time and sampling. It introduces a new approach to the analysis and design of relay-type control systems employing sampling. Because of the sampler, clamper, and relay, the system is nonlinear, and discontinuous in time. Systems of this general nature have been analyzed in the past by Chow[1] and Russell[2] using the sinusoidal response method introduced by Goldfarb[3] and Kochenburger.[4]

The investigation herein approaches the problem of analysis and design of relay-type feedback control systems with dead time and sampling in the time domain through the use of the phase plane method. By means of the phase plane, relay-type control action coupled with sampling and holding action can be considered as a varying dead time in the time domain. In the phase plane there exists a limit band within which all possible limit cycles lie. The exact boundaries of the limit band can be predicted. Control of this band makes possible control of the maximum and minimum amplitude and the period of oscillation of the output variable. The investigation brings out the effect of minor loop feedback on the limit band. It will become evident that minor loop feedback permits tight control of the limit band. The phase plane approach provides an intuitive insight upon the effect on system performance of varying any one or a combination of the system parameters.

Description of System

The block diagram of the system under consideration in this investigation is shown in Fig. 1. Both system input $r(t)$ and output $c(t)$ are continuous functions of time. The sampler, designated by the switch, consists of a constant rate-sampling device plus holding circuit of such a nature that its output is a stair step, or so-called jump function. The power amplifier, or relay, has the characteristic shown in Fig. 1, where $\varepsilon^*(t)$, the sampler output, is the relay input, and $\pm k$ the relay output. The symbol Δ is termed the relay hysteresis or differential gap. The output of the relay is a series of rectangular pulses of constant amplitude and varying period. T_d in Fig. 1 represents a fixed value of dead time.

With reference to Fig. 1, the following equations are written:

$$\varepsilon(t) = r(t) - [\alpha m(t) + c(t)] \tag{1}$$

$$T\dot{c}(t) + c(t) = K_3 m(t) \tag{2}$$

where $\dot{c}(t)$ denotes the time derivative of $c(t)$.

$$m(t) = \begin{cases} kK_1 \text{ if } \varepsilon^*(t-T_d) > \Delta \\ \pm kK_1 \text{ if } \Delta > \varepsilon^*(t-T_d) > -\Delta \\ -kK_1 \text{ if } \varepsilon^*(t-T_d) < -\Delta \end{cases} \tag{3}$$

From equations 2 and 3

$$T\dot{c}(t) + c(t) = \begin{cases} kK_1K_3 \text{ if } \varepsilon^*(t-T_d) > \Delta \\ \pm kK_1K_3 \text{ if } \Delta > \varepsilon^*(t-T_d) > -\Delta \\ -kK_1K_3 \text{ if } \varepsilon^*(t-T_d) < -\Delta \end{cases} \tag{4}$$

Substituting equation 2 into equation 1 yields

$$\varepsilon(t) = r(t) - \frac{\alpha}{K_3}[T\dot{c}(t) + c(t)] - c(t) \tag{5}$$

Let

$$c(t) = r(t) - \xi(t) \tag{6}$$

Putting equation 6 into equation 4 yields

$$T\dot{\xi}(t) + \xi(t) = T\dot{r}(t) + r(t) + \begin{cases} -kK_1K_3 \text{ if } \varepsilon^*(t-T_d) > \Delta \\ \pm kK_1K_3 \text{ if } \Delta > \varepsilon^*(t-T_d) > -\Delta \\ kK_1K_3 \text{ if } \varepsilon^*(t-T_d) < -\Delta \end{cases} \tag{7}$$

and equation 5 becomes

$$\varepsilon(t) = -\left[\frac{\alpha}{K_3} r(t) + \frac{\alpha}{K_3} T\dot{r}(t)\right] + \left(1 + \frac{\alpha}{K_3}\right)\xi(t) + \frac{\alpha T}{K_3}\dot{\xi}(t) \tag{8}$$

If it is assumed that $r(t) = 0$, equations 7 and 8 become respectively

$$T\dot{\xi}(t) + \xi(t) = \begin{cases} -kK_1K_3 \text{ if } \varepsilon^*(t-T_d) > \Delta \\ \pm kK_1K_3 \text{ if } \Delta > \varepsilon^*(t-T_d) > -\Delta \\ kK_1K_3 \text{ if } \varepsilon^*(t-T_d) < -\Delta \end{cases} \tag{9}$$

$$\varepsilon(t) = \left(1 + \frac{\alpha}{K_3}\right)\xi(t) + \frac{\alpha T}{K_3}\dot{\xi}(t) \tag{10}$$

A similar procedure may be followed for the case when $r(t)$ is constant.

To normalize equations 9 and 10 let

$$\xi = TkK_1K_3X$$

$$t = T\tau$$

Then equations 9 and 10 become

$$x''(\tau) + x'(\tau) = \begin{cases} -1 \text{ if } \eta^*\left(\tau - \frac{T_d}{T}\right) > a \\ \pm 1 \text{ if } \eta^*\left(\tau - \frac{T_d}{T}\right) > a > -\eta^*\left(\tau - \frac{T_d}{T}\right) \\ 1 \text{ if } \eta^*\left(\tau - \frac{T_d}{T}\right) < -a \end{cases} \tag{11}$$

$$\eta^*(\tau) = \lfloor x(\tau) + hx'(\tau) \rfloor \tag{12}$$

where the symbol $\lfloor\ \rfloor$ is used to denote the stair-step-type function.

$$h = \frac{\alpha}{\alpha + K_3}$$

and

$$a = \frac{\Delta}{(\alpha + K_3)TkK_1}$$

Considering the relationship between on-off control and sampling action, equations 11 and 12 are rewritten as

$$x''(\tau) + x'(\tau) = \begin{cases} -1 \text{ if } \eta(\tau - \lambda) > a \\ \pm 1 \text{ if } |\eta(\tau - \lambda)| < a \\ 1 \text{ if } \eta(\tau - \lambda) < -a \end{cases} \tag{13}$$

$$\eta(\tau) = x(\tau) + hx'(\tau) \tag{14}$$

where λ is a stochastic parameter varying between T_d/T and $(T_d + L)/T$. L is the sampling period.

Phase Plane Analysis

The phase plane is a plot of the system error x as a function of error rate x'. Because of the on-off action of the relay, the system under investigation will have two modes of operation, one when the relay

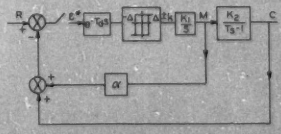


Fig. 1. System diagram

Paper 58-1026, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958, and re-presented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted May 7, 1958; made available for printing October 22, 1958.

KEISUKE IZAWA is with Purdue University, Lafayette, Ind., and LYNN E. WEAVER is with the University of Arizona, Tucson, Ariz.

This work was done under the sponsorship of the Argonne National Laboratory.

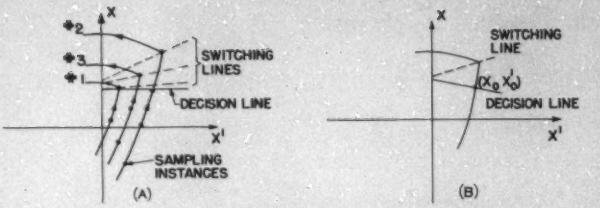


Fig. 2. A—Phase plane trajectories showing effects of sampling. B—Portion of phase plane trajectory showing switching and decision lines

output is $+1$ and the other when it is -1. The phase plane will thus be divided into two parts with relay operations occurring on the dividing line between the two regions.

For relay-type control systems without sampling, the dividing line is fixed. Introduction of the sampler and clamper causes this line to vary, the variation being caused by the varying dead time λ introduced by the sampler and clamper in conjunction with on-off control action. To clarify this, consider Fig. 2(A) showing a portion of the phase plane trajectories, where the dots on the trajectories represent sampling instances. The solid line, termed the decision line, is the division between the on and off control modes in the absence of sampling corresponding to $\eta(\tau)=\pm a$ of equation 14. The dashed line, called the switching line, corresponds to the dividing line between on and off control modes in the presence of sampling corresponding to $\eta(\tau-\lambda)=\pm a$ of equation 14. To illustrate the varying dead time concept, first consider trajectory 1. This is indicative of the case when the the sampler samples at the instant the phase trajectory is slightly past the decision line. At that instant, relay action occurs and the system switches to the other mode of operation. Trajectory 2 corresponds to the occurrence of sampling when the trajectory is slightly below the decision line. Since the sampling point falls below the decision line, relay action does not occur and the system must wait one sample period before switching modes of operation. Trajectory 3 is an example of any arbitrary switching. The switching line has two extreme limits, one limit when sampling occurs on the decision line, in which case it becomes the switching line associated with $\lambda=T_d/T$, the other limit when sampling takes place the instant before the trajectory crosses the decision line, in which case $\lambda=(T_d+L)/T$. It follows therefore that the effect of sampling on the phase plane plot is to cause the switching line to vary. The location of the varying switching line in the phase plane may be derived as follows.

Considering equation 13

$$x''+x'=\pm 1$$

Solving equation 13 for x, where x_0 and x_0' are the initial conditions, gives x_0 and x_0'.

$$x=x_0+(1-\epsilon^{-\tau})x_0'\pm(\epsilon^{-\tau}+\tau-1) \tag{15}$$

Differentiating equation 15 and solving for x_0' yields

$$x_0'=\epsilon^{\tau}x'\pm(\epsilon^{\tau}-1) \tag{16}$$

Substituting equation 16 into equation 15 yields

$$x_0=x-(\epsilon^{\tau}-1)x'\pm(\epsilon^{\tau}-1-\tau) \tag{17}$$

Since the relay will trip the moment the error signal η becomes greater than $+a$ or less than $-a$, the decision lines are expressed by

$$\eta=x+hx'=\pm a \tag{18}$$

Now, x_0 and x_0', the initial values of x and x', lie on the decision line; see Fig. 2(B). Therefore the equation of the switching line is given by substituting x_0 and x_0' of equations 16 and 17 for x and x' in equation 18, and is obtained as

$$x=(\epsilon^{\lambda}-1-he^{\lambda})x'\pm a\pm(\epsilon^{\lambda}-1)(h-1)\pm\lambda \tag{19}$$

where λ is the varying dead time.

The two switching line equations are for on-off

$$x=(\epsilon^{\lambda}-1-he^{\lambda})x'+a+(\epsilon^{\lambda}-1)(h-1)+\lambda \tag{20}$$

and for off-on

$$x=(\epsilon^{\lambda}-1-he^{\lambda})x'-a-(\epsilon^{\lambda}-1)(h-1)-\lambda \tag{21}$$

From equations 20 and 21, it is obvious that varying λ will cause the slope and the x-axis intercept of the switching line to vary.

Limit Band

Observing the switching line equations 20 and 21, it is evident that there exist two extreme locations for the on-to-off and off-to-on switching lines, corresponding to the two extreme values of λ, i.e., $\lambda=T_d/T$ and $(T_d+L)/T$.

Fig. 3(A) illustrates the two extreme limit cycles corresponding to the two extreme values of λ. Since λ varies in a stochastic manner between T_d/T and $(T_d+L)/T$, all steady-state trajectories will lie between the two limit cycles of Fig. 3(A). The band formed by the two limit cycles is called the limit band. Once any trajectory enters the limit band, it cannot escape from this band. The inner boundary of the limit band gives the minimum possible amplitude and period of the steady-state trajectories. In like manner the outer boundary of the limit band gives the maximum possible amplitude and period of the steady-state trajectories.

Considering the system equation corresponding to the off cycle

$$x''+x'=-1 \tag{22}$$

equation 23 is obtained with the initial values x_0 and x_0'

$$-x=x'-x_0'-x_0+\ln(x_0'+1)-\ln(x'+1) \tag{23}$$

Equation 23 describes the path of the

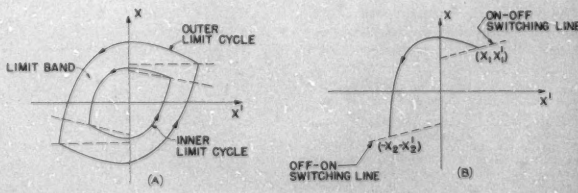
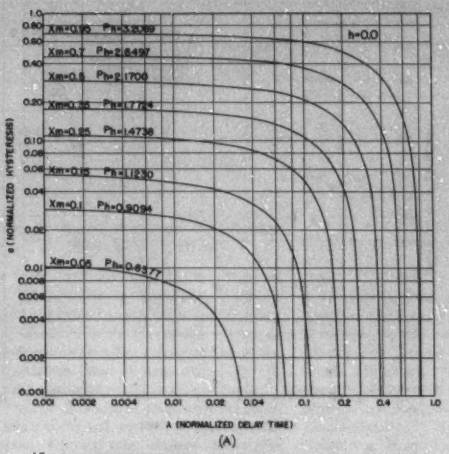
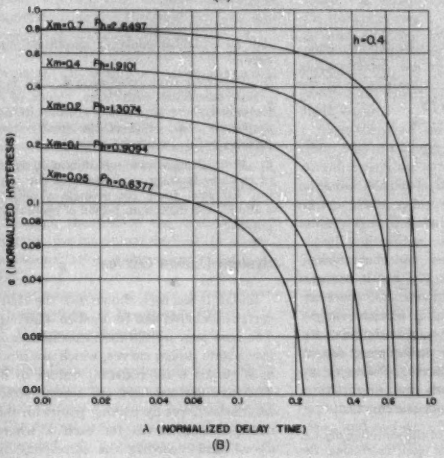


Fig. 3. Phase plane plot

A—Showing limit band
B—Off-trajectory for outer or inner limit cycle





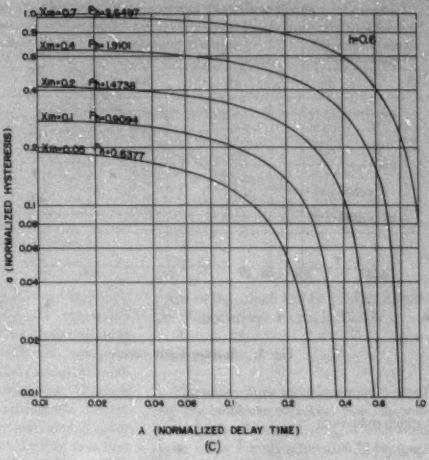


Fig. 4. System characteristic curves

A—For h = 0
B—For h = 0.4
C—For h = 0.6

phase plane trajectory corresponding to the off cycle of the relay. Consider the phase plane trajectory in Fig. 3(B) described by equation 23. Let this trajectory be representative of the two extreme trajectories corresponding to the boundaries of the limit band. Then (x_1, x_1') is a point on the on-off switching line, and $(-x_2, -x_2')$ a point on the off-on switching line corresponding to the inner or outer limit cycle; (see Fig. 3(A). Fig. 3(B) represents a portion of the outer or inner limit cycle, and the following two equations are formed:

$$x_2 = -x_2' - x_0' - x_0 - \ln(-x_2'+1) + \ln(x_0'+1) \quad (24)$$

$$-x_1 = x_1' - x_0' - x_0 - \ln(x_1'+1) + \ln(x_0'+1) \quad (25)$$

Subtracting equation 24 from equation 25 yields

$$x_2 + x_1 = -x_1' - x_2' + \ln(x_1'+1) - \ln(x_2'-1) \quad (26)$$

Because of symmetricity of the two pairs of extreme switching lines, $x_1 = x_2$ and $x_1' = x_2'$ in Fig. 3(B).

Replacing x_1 and x_2 by x, and x_1' and x_2' by x' in equation 26, equation 27 follows:

$$x = -x' + \frac{1}{2}\ln\left(\frac{1+x'}{1-x'}\right) \quad (27)$$

From equation 27 x and x' correspond to the values given by the switching line equations 20 and 21 for the two extreme values of λ. Substitute for x in equation 27 its value as given by equation 20. This substitution results in

$$a = (e^{\lambda}-1)(1-h) - \lambda - e^{\lambda}(1-h)x' + \frac{1}{2}\ln\left(\frac{1+x'}{1-x'}\right) \quad (28)$$

where $\lambda = T_d/T$ or $(T_d+L)/T$.

To find x' in terms of x_m, the possible maximum and minimum values of x corresponding to the outer and inner limit cycle, set $x' = 0$ in equation 23.

$$x_m = x' + x - \ln(x'+1) \quad (29)$$

Since, from the switching line equation

$$x = (e^{\lambda}-1-he^{\lambda})x' + a + (e^{\lambda}-1)(h-1) + \lambda \quad (30)$$

combining equations 29 and 30 yields

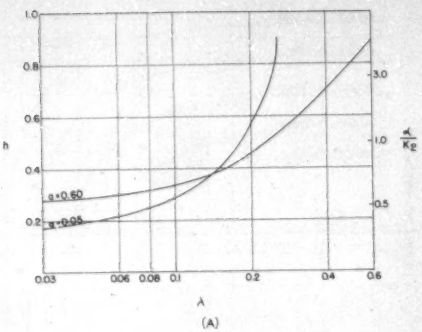
$$x_m = e^{\lambda}(1-h)x' - \ln(x'+1) + a + (e^{\lambda}-1)(h-1) + \lambda \quad (31)$$

From equations 27 and 30

$$\frac{1}{2}\ln\left(\frac{1+x'}{1-x'}\right) = e^{\lambda}(1-h)x' - (e^{\lambda}-1)(1-h) + \lambda + a \quad (32)$$

Combining equations 31 and 32 yields

$$x_m = \frac{1}{2}\ln\left(\frac{1+x'}{1-x'}\right) - \ln(1+x') \quad (33)$$



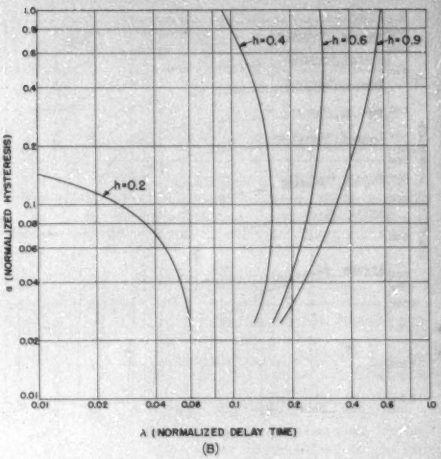


Fig. 5. System design curves

which reduces to

$$x' = \sqrt{1-\epsilon^{-3x_m}} \tag{34}$$

Equation 34 is substituted in equation 28, and the following expression is obtained:

$$a = (\epsilon^{\lambda}-1)(1-h) - \lambda - \epsilon^{\lambda}(1-h) \times \sqrt{1-\epsilon^{-3x_m}} + \frac{1}{2}\ln\frac{1+\sqrt{1-\epsilon^{-3x_m}}}{1-\sqrt{1-\epsilon^{-3x_m}}} \tag{35}$$

where $\lambda = T_d/T$ or $(T_d+L)/T$.

Equation 31 shows the relationship between the possible maximum and minimum value of the normalized output variable and the normalized system parameters. When the value of the system parameters are known, the maximum value of the output variable can be predicted. To evaluate equation 35, a digital computer was used. For a fixed value of h, incremental values of x_m were chosen. Values of a, corresponding to a series of values of λ, were calculated for each x_m; h was then changed and another set of calculations made. From the computer evaluation, a set of tables were constructed for values of $h=0$, $h=0.2$, $h=0.4$, $h=0.6$, $h=0.8$, and $h=0.9$. To visualize better the relations between system parameters and the output variables, a family of curves was plotted for the various values of h see Fig. 4. From these curves, the possible maximum and minimum values of the amplitude of the steady-state oscillations of the output variable can be predicted. The half-period of any symmetrical limit cycle is uniquely expressed by

$$P_h = \ln\left(\frac{1+x'}{1-x'}\right) \tag{36}$$

where x' and $-x'$ are the abscissas of the two symmetrical switching points.

If equation 34 is substituted into equation 36, the half-period corresponding to the outer or inner limit cycle of the limit band is given by

$$P_h = \ln\left[\frac{1+\sqrt{1-\epsilon^{-3x_m}}}{1-\sqrt{1-\epsilon^{-3x_m}}}\right] \tag{37}$$

With equation 37, for each value of x_m a corresponding value of P_h can be calculated and included in the system characteristic curves; see Fig. 4.

Several examples of characteristic curves are shown in Fig. 4. It becomes evident, upon examination of the characteristic curves, that an intuitive insight on the effect of parameter variation of the system steady-state performance can be obtained from the curves. Following are six conclusions drawn from observations of the system characteristic curves.

1. For a fixed value of the sampling period, delay and differential gap, increasing the minor loop feedback gain decreases the amplitude and increases the frequency of the steady state oscillations of the output variable.

2. For a fixed value of feedback gain, differential gap, and delay, increasing the sampling rate decreases the amplitude and increases the frequency of the steady-state oscillations.

3. For a fixed value of feedback gain, delay, and sampling period, increasing the differential gap of the relay will cause an increase in the amplitude and period of the steady-state oscillations.

4. There is a region associated with small values of λ for which the amplitude and period of the steady-state oscillations remains essentially constant. This implies that for certain sampling rates the output will have almost constant amplitude and period.

5. Following from conclusion 4 there is a region associated with large values of λ for which there is a wide variation in the amplitude and period of the steady-state oscillations of the output variable.

6. For each maximum and minimum value of the steady-state oscillations of the output variable, there corresponds a unique value for the maximum period of oscillation possible.

System Design Curves

So far, it has been shown how the characteristic curves can be used in analysis. The next step will be the introduction of the system design curves, which are plots of h versus λ for constant values of a. They are derived from the system characteristic curves by picking points on the characteristic curves for each h where $a=x_m$, and recording the corresponding value of λ. Through the use of system design curves, shown in Fig. 5(A) the feedback gain, or the varying dead time, can be selected so that the oscillations of the output variable will lie within a region bounded by $\pm a$. In other words, a feedback gain, sampling rate, or fixed dead time can be chosen so that the amplitude of oscillation of the output variable will always be less than a. To illustrate the use of the design curves, suppose the following specifications were placed on the system shown in Fig. 1:

Fixed dead time = 0.1
Sampling period = 0.05
Hysteresis = 0.1
$x_m \leq a = 0.1$

Fig. 6. Approximate shape of probability distribution curve for x_m

Table I. Comparison of Computer and Theoretical Results

Runs	Normalized Parameter Values h	a	Measured Maximum Values of x_m, P_h x_m	P_h	Expected Maximum Values of x_m, P_h x_m	P_h
1	0	0.03	0.34	2.0	0.34	2.0
2	0	0.3	0.79	3.0	0.80	3.0
3	0.2	0.1	0.36	1.9	0.36	1.9
4	0.2	0.3	0.62	2.6	0.62	2.6
5	0.4	0.2	0.36	2.0	0.35	1.8
6	0.4	0.6	0.79	3.0	0.80	3.0
7	0.6	0.2	0.23	1.5	0.24	1.5
8	0.6	0.4	0.43	2.2	0.43	2.0
9	0.8	0.3	0.28	1.7	0.28	1.6
10	0.8	0.6	0.55	2.5	0.55	2.3

For all runs $\lambda_{max} = 0.075$, and $\lambda_{min} = 0.325$.

The amount of minor loop feedback gain required to meet these specifications is obtained from the design curves as follows: Draw a line corresponding to the maximum value of dead time λ_{max}. In this example, $\lambda_{max} = 0.1 + 0.05 = 0.15$. Find its intersection point with the specified a curve and read the value of $h = 0.38$ from the ordinate. Points above the constant a curve correspond to $x_m < a$, and points below the curve to $x_m > a$.

Another set of design curves constructed from the characteristic curves is shown in Fig. 5(B). These curves are a plot of a versus λ for values of $x_m = a$ and constant h. This set of curves is used to select λ_{max} so that $x_m < a$ when the minor loop feedback gain is fixed. For example, if the (a, λ_{max}) point falls above the constant h curve, then $x_m < a$ if it falls below $x_m > a$. Through the use of the system design curves, the maximum limit of the steady-state oscillation of the output variable can be preset.

Comparison of Predicted and Computer Results

Simulating the system shown in Fig. 1, two analog computer runs were made for each value of $h = 0$, $h = 0.2$, $h = 0.6$, $h = 0.8$. The system output and output rate were recorded on a Sanborn recorder. To illustrate the limit band concept, the steady-state phase trajectories were displayed on a cathode-ray oscilloscope and a picture taken for several of the runs made. These clearly indicate the limit band and varying limit cycles. The maximum value of the output variable was obtained from these recordings and normalized. The maximum half-period was also taken from the recordings and normalized. These normalized values were then compared with the predicted values of x_m and P_h obtained from the system characteristic tables and curves. Table I compares the analog computer results with those obtained from the characteristic curves.

In comparing the theoretical and computer results, it was noticed that the maximum possible value of the output variable checks very closely with the computer result. However, in comparing the minimum possible value with that obtained from the computer, a considerable discrepancy was noticed. The minimum possible value was always considerably less than the computer value. This difference may be resolved in the following manner: Consider a portion of the phase plane plot shown in Fig. 6. At steady state the phase plane trajectory is always within the limit band.

Since the switching from on-state to off-state may occur only within the shaded area between the extreme switching lines and the boundary of the limit band, it must be at least proper to assume that switching in the shaded area occurs with equal probability. Now, when x_m, the maximum value of the each trajectory crossing x-axis, is considered, the crossing between maximum and minimum possible value of x_m will no longer occur with equal probability. The probability $P(x)dx$ with which x_m occurs between x_m and $x_m + dx$ is proportional to the area of the strip segment cut from the shaded area by two parallel trajectories which cross the x-axis at x_m and $x_m + dx$. A rough sketch of the probability density $P(x)$ is shown in Fig. 6. For the cases tested by the analog computer, the width of the limit band is relatively larger than the distance between the two extreme switching lines. It is found that the peak value of $P(x)$ occurs at a value of x closer to the maximum possible x_m.

The same discussion may be applied to the other switching region. Therefore, the succeeding switching point in the shaded area no longer occurs with equal probability, but will have zero probability along the outer and inner limit cycle and a maximum probability in between. This indicates that the maximum and minimum of x_m always lie within the maximum and minimum possible value of x_m, which can be predicted from the characteristic curves in Fig. 4.

Conclusions

The phase plane technique may be used for the analysis of relay type, second-order feedback control systems with dead time. The sampling function in conjunction with the type of system considered here may be treated as a stochastically varying dead time.

For systems of the type considered in this investigation, there exists in the phase plane a limit band within which all steady-state trajectories must lie. The limits of this band can be predicted from the phase plane analysis.

References

1. Contactor Servomechanisms Employing Sampled Data, C. K. Chow. *AIEE Transactions*, vol. 73, pt. II, March 1954, pp. 51–64.

2. Design Criterion for Stability of Sampled-Data On-Off Servomechanisms, Frederick A. Russell. *Thesis*, Columbia University, New York, N. Y., June 1953.

3. On Some Non-Linear Phenomena in Regulatory Systems (in Russian), L. C. Goldfarb. *Avtomatika i Telemechanika*, Moscow, U.S.S.R., vol. 8, no. 5, Sept.-Oct. 1947, pp. 349–83. (English Translation: *Report 1691*, United States National Bureau of Standards, Washington, D. C., May 29, 1952.)

4. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, Ralph J. Kochenburger. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

Discussion

J. E. Gibson (Purdue University, Lafayette, Ind.): The introduction of stochastic parameters into the analysis of a sampled data system with a relay is one of the most interesting points of this paper. The use of the stochastic parameter λ and the establishment of the boundaries within which a limit cycle must exist appears to be a very useful approach. Apparently the authors have established the necessary conditions for the existence of this limit band to their own satisfaction and indeed it is not difficult to

demonstrate the upper and lower bounds by the use of the describing function analysis. However, this discusser feels that this proof should be demonstrated rather than merely mentioning in the simple statement preceding equation 22 that the limit band exists.

In the authors' comparison of predicted and computer results, the heuristic presentation of the shape of the probability density of the x-axis crossings shown in Fig. 6 seems reasonable but certainly the conclusions would be strengthed if this probability density were derived rigorously and then found to agree with the actual simulator runs. This discusser also feels that a plot of the data in Table I would simplify the task of comparing the predicted and simulator results.

The block diagram shown in Fig. 2(A) may be rearranged into a single loop diagram consisting of a sampler plus hold, the relay, and a linear transfer,

$$H=\frac{aK_1(hTS+1)}{hS(TS+1)}$$

This fact will allow the system design curves shown in Figs. 4 and 5 to be employed for a single-loop system with an integration and with or without a lag or a lead.

Finally, this discusser wishes to raise the question of the system response to various inputs, other than the step function mentioned following equation 10. He realizes that this is an easy question to ask and an extremely difficult one to answer in any generality. However, any data the authors might have obtained previously or since their paper was written would make a welcome addition to their already quite useful work.

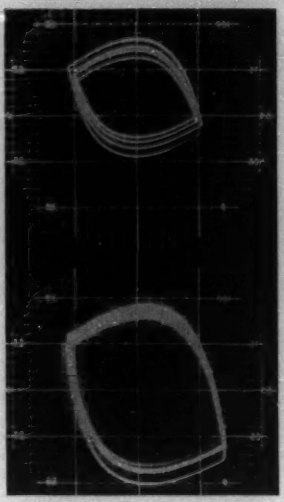


Fig. 7. Limit band and varying limit cycle; see Table I

Keisuke Izawa and Lynn E. Weaver: The authors thank Prof. Gibson for his comments. To verify the existence of the limit band, consider equations 20, 21, and 27. Actual switching between two trajectories in the phase plane, i.e., the on and off trajectories, occurs λ after the corresponding decision is made (on-off at $x=a$ and off-on at $x=-a$). This switching will take place along straight lines given by equations 20 and 21. If the switching point, which is indicated by equations 20 and 21, satisfies equation 27, then a periodic limit cycle exists. Since λ varies between $\lambda_{min}=T_d/T$ and $\lambda_{max}=(T_d+L/T)$, two possible corresponding symmetrical limit cycles exist, and form the boundaries of the limit band shown in Fig. 3(A). These two closed trajectories are the extreme cases in the steady state and may or may not actually occur, depending on whether the switching point corresponding to the extreme values of λ satisfies equation 27. Furthermore, since the time duration between two succeeding switching points must be an integral multiple of the sampling time L, the existing trajectories, after the transient, would always be within the limit band. In other words, if half period $P_{\lambda max}$ corresponding to λ_{max}, given by equations 28 and 36 is smaller than $(N+1)L$ and larger than NL, N being an integer, then the maximum possible half-period would be NL, for which the corresponding trajectory lies inside the outer boundary of the limit band. The same discussion follows for the minimum possible trajectory. A quantitative verification of the existence of the limit band and varying limit cycle is given in Fig. 7, which shows the cathode-ray oscilloscope display of the steady-state phase plane portrait corresponding to analog computer runs 3 and 4 as listed in Table I. Run 3 corresponds to the upper portrait and run 4 to the lower portrait.

The authors used the heuristic presentation of the shape of the probability density curve of the x-axis crossings to explain the discrepancy between the predicted steady-state minimum value of the output variable, and that actually obtained with the analog computer. The predicted maximum values agreed closely with the values actually obtained. In either case the steady-state value of the output variable fell with the limits set by the limit band. Since, in practice, interest will mainly be in predicting the maximum possible steady-state amplitude of the output variable, no further effort has been made to derive rigorously the probability density curve. Such a derivation however, would certainly strengthen our conclusions.

Prof. Gibson's rearrangement of Fig. 2(A) to a single-loop system with a lead term helps to extend further the technique presented in the paper. In answer to his final comment, work is currently being done to apply the techniques developed in this paper to other types of nonlinearities as well as various inputs.

Control of Infrared Radiation

IRA J. BARBER
MEMBER AIEE

THE CHARACTERISTICS of commonly used infrared sources have been rather well delineated in the past,[1,2] so that the quantity and quality of their radiant outputs need not be covered in this paper. Generally speaking, these sources have one or two dimensions, represented by point sources, line sources, and area sources. All are basically omnidirectional, and their utilization is dependent on some form of optical control, both individually and when combined to form an oven.

It is with the type of control, its optics, and the geometry of utilization of these units in multiple that this paper is chiefly concerned. Except for the specific reflectivity of the material involved, the spectral characteristics of the source are not of importance in consideration of these points. The immediate directional control of an infrared generator is important where operating expense must be considered, in practically all line production ovens. It is desirable to transform the largest possible share of the initially developed radiant energy into useful radiation (i.e., absorbed by the product) and the efficiency of an individual reflector is shown by the ratio of forward useful radiation to initially generated radiation. Prime factors influencing this amount are optical design of the reflector, and reflectivity of materials used.

In initiating a study of the most effective methods of distributional control of radiation, it is first necessary to

Paper 59-31, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 31, 1958; made available for printing November 19, 1958.

IRA J. BARBER is with The Fostoria Pressed Steel Corporation, Fostoria, Ohio.

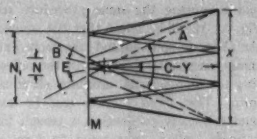


Fig. 1. Energy reaching target from bare source

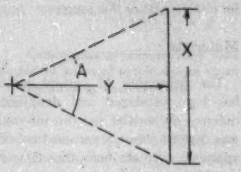


Fig. 2. Energy control through flat reflector

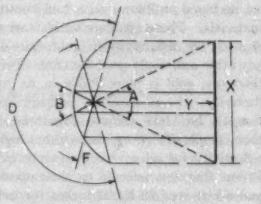


Fig. 3. Properly contoured reflector, which can greatly increase energy reaching target

establish the desired end result. Purposes for directed radiation will vary widely, and the solution for one problem may not be the most suitable for another. The most usual requirement is to provide a uniform field of radiant flux at some selected distance. To a lesser extent, there are requirements for concentrated lines and concentrated spots. This paper will deal with the development of specific configurations for each.

Optics of Individual Reflector

Considering initially, then, the point sources, represented principally by the globular tungsten filament lamp, the problem usually is to start with a concentrated source of radiation and through a direction and redirection of its rays provide a heating area of substantial uniformity at a given distance from the source. In addition, in industrial applications this uniform field should be produced by these units in multiple since the use for a single unit is very small. Reference to Fig. 1 will show the effect of the use of a flat reflector back of the source. Graphical appraisal of the relative intensities thus obtained will illustrate that the energy resulting on a plane at a distance y from the source will be about double the amount which will result from direct radiation with no reflector at all; see Fig. 2. If, however, a properly contoured reflector is provided for the source (Fig. 3), it is possible to increase this amount to four or five times that of the bare source. Thus, from the same unit input of energy, it is possible to provide effectively several times the original amount of radiation at a useful location at no increase in power consumption.

To examine this statement more closely, consider the condition of a bare source. Assume a product of x dimension of zero reflectivity at a distance of y from this source. Under these circumstances, only a percentage of the developed energy equal to angle $A/360$ (considered in a plane) will initially reach the product; see Fig. 2. Since none will be rereflected, it is obvious that this is the total amount of energy which can be expected to develop useful work.

Now, if a flat reflective surface m be placed at some point back of the source with the other conditions of distance and dimension held equal, it is apparent that the amount of energy reaching the surface x has been increased by the amount of the angle $B/360$ times the reflectivity of m; see Fig. 1. The total energy reaching x would then become $A/360+BR_m/360$. For constant values of x and y, it is apparent the angle B decreases as the distance back of the source is increased. This can be of importance where x and y are small; it is of little or no value as these dimensions approach infinity. The small dimension n is all that is necessary where the reflectivity of x is zero. If the reflectivity of x (R_x) is greater than zero, the larger value of n, becomes involved. As the dimension x increases, with $R_x > 0$ and with multiple sources, the need for a continuous flat reflector becomes apparent. If, however, $R_x = 0$, no reflector beyond n is required. A rule may be stated, then:

The importance of net oven reflectivity increases with the reflectivity of the product.

Returning to Fig. 1, if $R_x > 0$, the total energy absorbed by x would be, roughly

$$a_x/360[(A-C)+(B-E)R_m+Ca_xR_m+Ea_xR_m^2]$$

If a parabolic shape p is introduced in such a fashion that its total face width

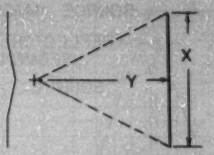


Fig. 4. Improperly contoured reflector, which may add nothing to energy on product

is equal to x and the source is in its focal point, the amount of energy reaching x in addition to that represented by $A/360$ would be $DR_p/360$. If $R_x > 0$, this flux represented by $D/360$ would be reflected and rereflected between the parabolic surface p and the surface x until it divides itself so that the energy from the reflector is

$$a_x/360\{(D-2F-B)R_p+(B+2F)/[(1-R_p)+a_x]\}$$

and the total energy resulting in x equals

$$a_x/360\{A+(D-2F-B)R_p+(B+2F)/[(1-R_p)+a_x]\}$$

both equations approximate.

If specific values are assigned to the foregoing symbols (angle $A = 60$ degrees, $a_x = 100\%$), the following rough comparison would exist. In the case of no reflector, 60/360, or 16.7% of the radiant energy developed by the source would reach x. If a flat surface m of high reflectivity were used at a distance of $0.1x$ in back of the source where $R_x = 0$, the resultant reflected angle B would approximately equal A and the additional amount of energy reaching x would be 16.7% times the assumed 95% value of R_m, or 15.8%. Thus, the efficiency of transfer would have been increased from 16.7 to 32.5%, or nearly doubled.

In the event of the third case in which a contoured reflector is employed, the initial direct radiation again = 16.7%. The value of the reflected energy provided through the large angle D, however, is now nearly 54% and the total percentage of initially developed energy in forward useful radiation over 70%, or something more than four times that achieved where no reflector is involved and over double that which exists from a flat oven wall of equally reflective material and optically unadapted to the source location. It should be noted in passing that it is possible (Fig. 4) to illustrate a condition in which a reflector adds nothing to energy in a given area. In this case, the forward useful radiation remains 16.7%. While this would rarely be done deliberately, it is obvious that

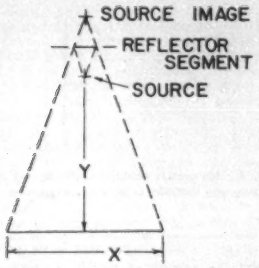


Fig. 5. Line from source image to limits of work area, which marks dimension of reflector segment

the use of a corrugated backing sheet with indiscriminate source locations would inevitably lead to some degree of this condition.

If the dimension x is made infinite, the percentage of impinged energy with no reflector reaches a maximum just under 50%. In the case of the flat reflecting surface, the same reasoning holds since the angle of lost energy approaches zero and the utilization is just less than theoretical maximum of 97% which is, however, double that of no reflector. Where contoured reflector p is used with x infinite, the theoretical maximum is 97½% where the reflectivity of m and p is 95%. To the extent that the flat oven walls are not continuously reflective but are composed partly of areas of no reflectivity (zero per cent) or poorly reflective materials, the net oven radiant efficiency will vary between 50% and 100% of that which could be achieved with continuous flat reflective walls. Similarly, to the extent that a material of lesser reflectivity than 95% is introduced as a reflector material, the maximum net reflectivity of the infinitely large oven will vary from about 50% to 100% of the ultimate.

In the mention of a reflector optically contoured to fit a job, reference was made to a parabolic form. While this is frequently a satisfactory solution, where the source is located properly with respect to its focal point, for various reasons it is frequently desirable to provide an equal control through the use of reflectors composed of properly angled flat surfaces rather than continuous curves. In this latter instance, the development of the desired surface geometry through the use of source images is illustrated in Figs. 5 and 6.

Control for area sources is practically restricted to flat reflectors. In the case of no reflectors, the percentage of radiation forward and backward from a heated plate will be a function of the relative emissivities of the two surfaces. Where these are equal, it is obvious that 50% in one direction will be a maximum. The per cent of radiation in one direction may be determined by dividing the emissivity of that side by the sum of the two emissivities. In determining actual figures, where a flat backing-up reflective surface, is employed, and if the plate is transparent or translucent, the relative surface temperatures must be known as well as the opacity of the coating.

If concentration into a spot or line of heat is desired, it is necessary only to define the area desired to be covered, and to pursue the same procedure see Fig. 7. It will be noted that the smaller the ultimate spot desired, the more and smaller segments there are in the reflector, to approach an ultimate elliptical form. This cross section can be extended linearly to develop a trough for a linear source, or rotated to form a round reflector for a point source. It will be seen that the same basic rules would apply to development of reflectors for either point or line sources.

Materials

The reflectance of various materials has been cataloged and described by different authorities and this information was derived through various methods of appraisal. On the basis that the matter should be re-explored with the use of the total radiation theory dealt with later in this paper, a stable system was set up based on known areas and known materials. There are more authoritative data on the reflectivity of gold than of any other commonly used material. Furthermore, gold is more stable in its reflectivity at all usable temperature ranges,[3] has less surface deterioration than comparable materials,[4] and has high reflectivity for all source temperatures.[5] It was therefore selected as a standard, and a test area 22 by 22 inches was set up for the comparison. Holding the voltage constant and with the room ambient constant, sample specimens 22 by 22 inches of various materials were interchanged in the test area and their relative temperature rise noted. These temperature rises were assigned intensity valuations based on the curve in Fig. 8

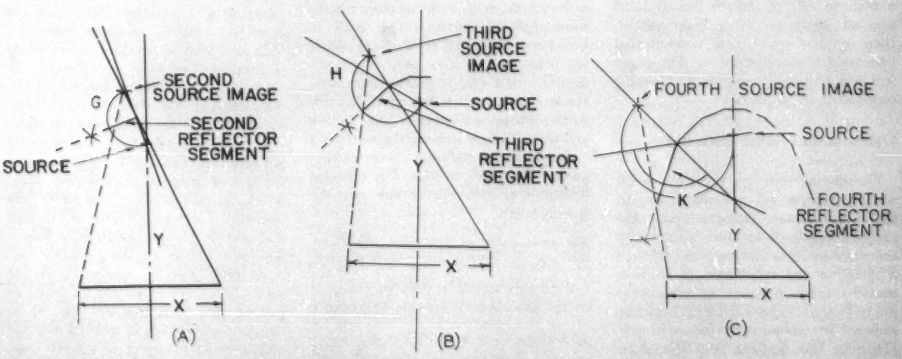


Fig. 6. Slopes fixed by bisecting lines G, H, and K, lines from source images to limits of work areas mark dimensions of reflector segments

A—Second image and segment B—Third image and segment C—Fourth image and segment

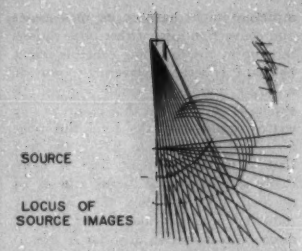


Fig. 7. Construction of concentrating type of reflector by means of source images which require many small segments

which has been fairly well certified over years of checking and application. The percentage of subject area was checked by substituting metals in the primary reflecting areas of the same materials as those being tested. This permitted accurate determination of the factors in the total radiation formula, so that it becomes:

$$K_1(a_p)/(K_2+K_3+K_4+K_5)=H$$

(equivalent heat units, Fig. 8), or $Ka_p=H$ where the constants K_1, K_2, K_3, K_4, and K_5 represent fixed products of area and absorptivity of the various parts of the test system. On this basis, the following values for specific reflectivity of the materials tested based on radiation from a 2,500 degree K source, were as follows:

Material	Per Cent
Specular gold plate	98
Alzak	78
Electrolytically brightened anodized aluminum	75
Unbrightened anodized aluminum	59
Aluminized steel	55

Data published by Hottel[6] indicate base metal reflectivities as high as 96% for highly polished aluminum with some 91% to 89% for the oxides. Gold is given at 98.2% with nickel 89%. These figures, however, are for the reciprocal of the emissivities over the entire infrared range, and do not reflect the preponderance of energy in specific bands applying to commercial sources. For industrial application, the relative reflectances used should be related to the characteristics of the source. These values are considered the specific reflectances of the materials within the meaning of this paper. High values are also assigned silver and copper but it must be considered that commercially speaking these are not practical figures because of the transience of pure metal surfaces which is due to their tendency almost immediately to develop a thin oxide film. This is especially true of aluminum.[7] Of all of these, gold continues to present a relatively pure metal face for reflectivity and commercial comparisons must be based to a degree on the relative value of metallic gold compared with the oxides or other salts of alternate metals. Alzak and anodized aluminum are deliberately formed oxides, which accounts for their relative low reflectivities in the infrared range.

Specularity in a reflective surface is important in that it redirects a larger portion of the impinging beam in a situation in which the proportion of the beam which is diffused laterally or along the plane of the reflecting surface is largely lost through lateral "spill" in average installations.

Oven Geometry

The net radiant efficiency of an oven enclosure can be approximated by the following method. The absorptivity of the materials involved being known and their areas being calculable, it is possible through the theory of total radiation[8] to predict the approximate absorption of the product x as follows:

A_r = surface area of oven wall
A_p = surface area of product
A_o = area of no wall, oven ends, gaps, etc.
A_f = area of floor, all poorly reflective surfaces
a_r = absorptivity of oven wall material
a_p = absorptivity of product
$a_o = 100\%$
a_f = absorptivity of floor, poor surfaces
E = total developed radiation
$A_pa_p/(A_ra_r+A_pa_p+A_oa_o+A_fa_f)$ = per cent E in product

The percentage of this amount absorbed by the product will represent the net oven radiant utilization. This, therefore, provides a simple method for predicting relative results by various types of oven construction and material. Further, where a standard test oven is so constructed that the constants are all known figures other than the product, it is possible to determine the product absorptivity quickly. Similarly, by using sources of definite radiating characteristics on a selected subject, it is possible to determine accurately the surface characteristics of the subject. By working continuously with known factors and introducing one variable at a time, it is possible to determine correct values of specific absorptivity for various metals and materials. (See preceding section on materials.)

It will be apparent that color sensitivities can be accurately determined by this method. Failure to consider total relationship between radiator and product has led to some incorrect data in published color absorptance tables. What has been tabulated has largely been the relative temperature rises in systems of low net radiant efficiency, and the poorer the net radiant efficiency, the wider the color disparity.[1] The foregoing method provides a means of accurately expanding pilot or laboratory test installations into final recommendations. It crystallizes the "size factor" to permit a more logical approach to layout problems.

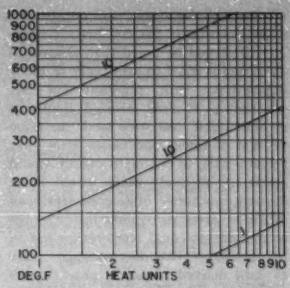


Fig. 8. Relationship between absorbed radiant energy and ultimate product temperature

To consider an example: A typical automotive oven is 10 feet wide by 10 feet high by 100 feet long. It has a total inside surface of 3,000 sq ft (square feet). It could contain, in operation, approximately five finished automobiles with a total exposed surface area of approximately 2,400 sq ft with an assumed absorptivity of 50%. There are approximately 1,000 sq ft of floor to this oven, estimated at 80% absorptivity, and there are 200 sq ft of ends. Assuming the oven walls to be highly reflective (5% absorption), the division of energy introduced in this oven would theoretically be: 3,000 by 0.5=150; 2,400 by 0.50=1,200; 1,000 by 0.80=800; 200 by 1.00=200. Thus, of a total of 2,350, only 1,200, or 51% of the input energy would be reflected in useful work. If, by contrast, the oven walls are painted black and the other factors remain the same, the calculation becomes: 3,000 by 1.00=3,000; 2,400 by 0.50 remains 1,200; 1,000 by 0.80 remains 800; 200 by 1.00 remains 200. Now, however, only 1,200/5,200, or 23% of the input energy, will be found in the product. Thus the reflectivity of the oven walls can account for more than doubling the

efficiency of the oven measured in terms of kilowatt-hours per car. (As a corollary, it should be noted that absence of oven wall is equal to oven walls of 100% absorptivity with resultant low net efficiency.) If plain oxidized aluminum or aluminized steel were used, the same method of calculation would show that roughly 34% of the input energy would reach the product, which would mean that approximately 51% more developed radiation would be required in the source to produce an equal temperature rise in the product. This is reasonably close to data derived in actual practice and serves to add validity to the theory. In the highly reflective oven first considered, operated without product, the radiated energy would theoretically divide itself as follows: 3,000 by 0.05=150; 1,000 by 0.80=800; 200 by 1.00=200; for a total of 1,150. The oven walls would thus receive 150/1,150, or 13% of the radiated energy instead of the 6.4% initially absorbed. If the oven walls had reached 150 degrees Fahrenheit in the first instance, they could now be expected to attain slightly over 200 degrees Fahrenheit. (The higher air temperature prevailing in a filled oven, however, tends to offset this effect, and in practice, little change can be measured.)

The foregoing material may be used to advantage in an examination of the relative efficacy of types of solution to an oven problem. For example, if it is desired to heat a panel of an assumed 2 ft by 2 ft with radiant energy, it is obvious that several oven selections can be made. It is possible to accomplish this result with one-sided radiation emanating from a radiating panel of the same size as the product. Assuming that this course is followed, the total radiation theory would indicate a reflector area of 4 sq ft with absorptivity of 5% (assuming reflectors to be gold), and a product area of 4 sq ft with an assumed absorptivity of 50%. If the distance between target and radiating panel is 6 inches, the area of total absorptivity (no reflecting walls) represented by ends, top and bottom, is 4 sq ft plus an equal area of total absorptivity of radiation from the back of the target. A total of 19.6% of the developed radiation would result in the product. If, however, two-sided radiation is employed, the radiating panels are increased to 3 feet by 3 feet, the unit is enclosed on two sides with reflective panels (as would be possible in a conveyorized oven) and, by means of a reflective silhouette, the end areas are reduced by providing an aperture for entry and departure with a 2-inch clearance in all directions, the flux can be recomputed as follows: 18 sq ft of reflecting area at (5%); 8 sq ft (at 50%) product area; an additional 10.44 sq ft (at 5%) (reflective baffling) with an area of total loss amounting to 1.56 sq ft of 100% absorptivity. On this basis 57.2% of the developed radiation (4 divided by 6.982) would end up in the product. Comparatively the total radiant energy required would be only one third that involved in the first solution; the amount of equipment required would be approximately five times as great and the rating of individual sources would be only 1/14 that initially considered. In practical terms, if the original solution had required four 1,400-watt sources, the latter solution would require 18 100-watt sources to perform exactly the same heating task. This illustrates a basic rule of oven geometry:

Oven-radiant efficiency varies directly with the percentage of surrounding area occupied by reflective material.

It is interesting to rework the foregoing example on the basis of a material such as aluminized steel with relatively low specific reflectivity. Here the figures will become 16.4% instead of the earlier 19.6% where the 2-ft by 2-ft radiant panel is employed and where the fully surrounded oven of the second example is constructed of the same low reflectivity material, the oven-radiant efficiency is increased only to 18.9%.

It should also be noted that if the total oven wall area is made up partly of highly reflective areas and partly of areas where the reflector is omitted, or of material of low reflectance, the resultant net oven efficiency is lowered in the same manner as if the reflecting areas were all of low reflectivity. This emphasizes the inherent error of oven constructions wherein blank or missing areas in oven walls are involved, or where consideration is given only to the reflectivity of individual reflectors, either separate or, as with reflector lamps, built into the source structure.

An oven of high radiant efficiency cannot be constructed of materials of low specific reflectivity.

A corollary rule may be derived from the foregoing: In ovens of low basic efficiency due to low percentage of enclosure area, selection of reflecting materials becomes of lesser importance. It should be remembered at this point that these conclusions deal with radiation only and do not take into consideration the effect that changes in physical oven structure might have on air movements and convection loss. Other observations relating to the foregoing example might be that it is better engineering to provide a larger number of small sources (18 versus 4) on the basis that failure of one source would drop the total flux only 5½% as compared to 25%; the fact that a uniform radiant field is more easily secured with a larger number of sources; and the fact that a larger radiating area provides better angular direction of radiation in so far as irregularities of product are concerned. Offsetting factors are the higher initial equipment costs as well as somewhat higher maintenance problems incident to increased equipment areas, and (perhaps) decreased accessibility. Oven geometry in the use of control equipments for point, line, or area sources is basically the same since all may be reduced to equal heated panels for this consideration. There is no critical angle of impingement in the application of radiant energy. Angular effects are subject wholly to the cosine aw. The proper engineering approach in all applications is, of course, to evaluate all factors specifically as a guide to final determination. There may be instances where intermittent use of equipment, mobility, or low ratio of operating cost to end value may dictate the deliberate sacrifice of operating efficiency in favor of low initial cost. It is, however, important that operating cost be properly weighed in full consideration of the problem. Many examples of economic comparison could be worked out to illustrate this point. A basic 10-year amortization working figure, which may surprise many, is:

One per cent saved in operating cost will usually justify 3% additional in first cost.

There are then three factors which can importantly affect the utilization of energy developed by any source applied in an oven structure. These may be recapitulated as follows: 1. Per-cent surround, or percentage of product surrounded by oven walls in terms of total enclosure. 2. Average specific reflectively of all interior oven materials. 3. Optical efficiency of the system both in individual reflectors and total oven. One striking conclusion appears in connection with the foregoing. In an average-sized oven with medium "fill," say 50%, where all the foregoing factors of design are omitted, it appears as though the total energy initially reaching the product could be as low as 10% of that developed by the source. The product absorptivity would permit utilization of no more than 80% of that amount

with perhaps 8% of the originally developed radiant energy thus ending in useful work. Where careful attention is paid to all of the foregoing factors and practical maximums are achieved, it appears likely that an ultimate of 70% to 75% might be realized. Thus, in oven design alone, with the same product, it would appear that over-all radiant efficiency might vary from 8% to 75%. Where the radiant efficiencies of various sources are taken into consideration (ranging from usually 45% to 87%), it will be seen that it is not difficult to visualize one infrared oven which is as much as 15 times as efficient with respect to radiant energy as another. While these extremes of deviation are not commonly encountered, the observation has been made that the difference between a good infrared oven and a bad one may be much greater than is usually possible in the case of convection ovens.

Summary

To recapitulate, the fundamental rules applying to the control and utilization of radiant heat sources, in single and in multiple housings which are in radiant-oven use may be stated as follows:

1. The radiant efficiency of an individual reflector is measured by the accuracy of its optical design and the specific reflectivity of its surface. It may attain a little more than 90%.

   Corollary A: A correctly contoured reflector may provide several times as much forward useful radiation as a flat reflecting surface, and many times as much as no reflector at all.

   Corollary B: An incorrectly designed reflector may contribute little or nothing to a specific heating problem.

   Corollary C: Any departure from specularity or "gloss" in a reflector surface detracts from its control efficiency, rather than its total reflectance.

2. The net reflective control of a radiant oven is measured by:

   A. The radiant efficiency of the individual reflectors involved.

   B. The per cent of total enclosure achieved.

   C. The net specific reflectivity of the materials making up the total oven interior.

3. The net thermal efficiency of a radiant oven is determined by:

   A. The radiant efficiency of the source.

   B. Its net reflective control.

   C. The absorptivity, effective area, and physical disposition of the product.

   D. The amount of convective loss involved.

References

1. Radiant Heating Using Lamps and Quartz Tubes, Ira J. Barber. *Proceedings*, National Industrial Electric Heating Conference, Cincinnati, Ohio, vol. 3, Feb. 4, 1958.

2. Industrial Application of Infrared, Ira J. Barber. *Electrical Engineering*, vol. 72, no. 9, Sept. 1953, pp. 764–69.

3. *WADC-TR 57-159*, Wright Air Development Center, Dayton, Ohio, 1957, p. 36.

4. Gold and Gold Alloys, E. M. Wise. "Corrosion Handbook," Electrochemical Society, Inc., John Wiley & Sons, Inc., New York, N. Y., 1948, p. 112.

5. Infrared Reflector Surfaces, J. D. Hall. "Industrial Applications of Infrared" (book), McGraw-Hill Book Company, Inc., New York, N. Y., 1947, p. 37.

6. Applied Heat Transmission (book), Herman J. Stoever. McGraw-Hill Book Company, Inc., 1941, p. 213, Table IV, H. C. Hottel.

7. Aluminum for Reflectors, Junius D. Edwards. *Transactions*, The Illuminating Engineering Society, New York, N. Y., vol. XXIX, no. 5, May 1934, pp. 351–57.

8. General Theory of Enclosure, H. H. Higbie. "Lighting Calculations" (book), John Wiley & Sons, Inc., 1934, pp. 249–51.

The Use of Frequency-Response Tests in the Analysis of a Foil Mill Automatic Gage Control

S. J. JONES
MEMBER AIEE

R. M. SILLS
MEMBER AIEE

FREQUENCY-response measurements are often made in the laboratory to determine performance of components or systems. Such tests are far more rare on complete processes. To the authors' knowledge, the tests described here were the first to vary sinusoidally, while rolling, the speed of a metal rolling mill to determine how strip thickness was affected by the frequency of oscillation.

The results obtained indicate that such tests can be useful for substantiating simplified theoretical analysis, determining quantitative relationships for complex transfer functions, and improving system performance.

Paper 59-78, recommended by the AIEE Metal Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing December 3, 1958.

S. J. Jones is with the Aluminum Company of America, Pittsburgh, Pa.; and R. M. Sills is with the General Electric Company, Schenectady, N. Y.

The authors are indebted to Messrs. R. G. Beadle and R. A. Phillips of the General Electric Company for the physical and mathematical concepts used in the Appendix.

Controlling Gage on a Foil Mill

Any strip less than 0.006 inch thick is classed as "foil," although aluminum strip rolled for containers of various sorts may be as heavy as 0.0105 inch but rolled on a "foil" mill. Although foil may be made of many metals, this paper is confined to aluminum foil, which will simply be referred to as "foil."

The process of cold rolling foil is quite different in many respects from those usually used for other products. Since this is not a treatise on foil rolling, only those characteristics that affect gage control will be mentioned.

Perhaps the most notable aspect is the fact that the rolls are always jammed together, or set "below face," such that roll pressure may be 100 tons or more per screw with no metal in the mill. This is also characteristic of tinplate steel mills on the last pass or two. This results in considerable roll flattening; so much so, in fact, that small changes in screwdown settings change only the deformation of the rolls and the cross-sectional shape of the foil, but not the thickness. This condition will be reached on any rolling mill when the ratio of roll diameter to metal thickness is made high enough.[1]

Another factor is the effect of mill speed on thickness. The "speed effect" occurring in all strip mills is well known, and is attributed largely to a decrease in the coefficient of friction between work rolls and strip as speed is increased.[1] On a foil mill, this effect is more pronounced than on mills rolling heavier thicknesses. In fact, it is so pronounced at all speeds at which foil is rolled that it is the means most often used by the operators to man-

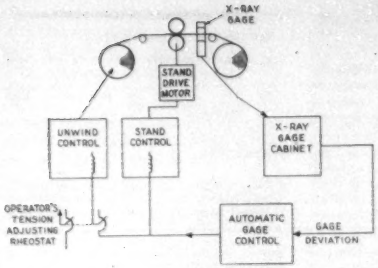


Fig. 1. Simplified schematic block diagram of original gage control system

ually adjust gage. It is also the method used in the gage control discussed in this paper.

Description of Gage Control

Fig. 1 is a schematic block diagram of the foil mill gage regulator, as originally designed, but with certain details omitted. An X-ray gage, mounted close to the work rolls, provides a thickness deviation signal. This is amplified by three stages of magnetic amplifiers, and the resultant voltage is applied to a control field on the mill voltage or speed regulator, to vary the speed in proportion to thickness deviation. The original design included the circuit shown to adjust unwind reel current (and hence tension) in proportion to deviation but this was later disconnected because it did not appear to be necessary. Other circuits included, but not shown, were forcing, which will be discussed later, and an integrating reset circuit to readjust the mill motor-operated rheostat to keep the regulator within range. This circuit does not enter into the analysis which is described here so it need not be considered further.

When the regulator was installed, it was found to be difficult to stabilize. Although the loop gain was not high, the system had to be made quite slow to obtain stability. Indications were that the gain increased in the middle frequency range. To learn how to improve its performance, it was decided to make frequency-response measurements on the complete system.

Preliminary Analysis

Prior to making the frequency-response tests, a preliminary block diagram analysis was made. Fig. 2 shows a somewhat abbreviated block diagram of the system. The Appendix gives the derivation of the $\Delta S_1/\Delta p_m$ block. In order to reduce the diagram to more manageable form it is necessary to make some simplifying assumptions.

As a start, it was assumed that $\Delta I/\Delta p_m = 0$. This states that the reel current regulators are fast compared with mill speed changes, and have a high gain, both of which seemed like good assumptions. This makes $\Delta \tau_d = 0$, and eliminates the reel voltage regulators, reel current regulators, and reel counter electromotive force regulators all at once, leaving only the tension loop ($\Delta S_1/\Delta p_m$) to be reduced as far as the reels are concerned.

The transfer function required is $\Delta S_1/\Delta p_m$. From the block diagram, using $C/R = G/1+GH$, the following is written

$$\frac{\Delta S_1}{\Delta p_m} = \frac{\dfrac{E_1' A_1}{p_1(1+pT_n)}}{1+\dfrac{E_1' A_1}{p_1(1+pT_n)}\dfrac{r_1^2}{pJ_1G_1^2}} = \frac{pT_{r1}E_1' A_1/p_1}{1+pT_{r1}+p^2T_{r1}T_n} \quad (1)$$

where $T_{r1} = J_1G_1^2p_1/E_1' A_1 r_1^2$, one of the time constants between torque and tension.

The system block diagram now can be redrawn as shown in Fig. 3.

It will be recalled that experience with the gage control indicated a rising gain in the middle frequencies. Since all elements of the control except the mill-strip-reels block were known, it could be assumed that any peculiar behavior originated in this area. It was the objective of the preliminary analysis, therefore, to determine a transfer function for the mill. This had to be simple enough to have physical significance and to be verified by tests.

With no further justification, then, other than that subsequent tests seemed to prove the assumptions reasonable, the block diagram of Fig. 3 was further simplified.

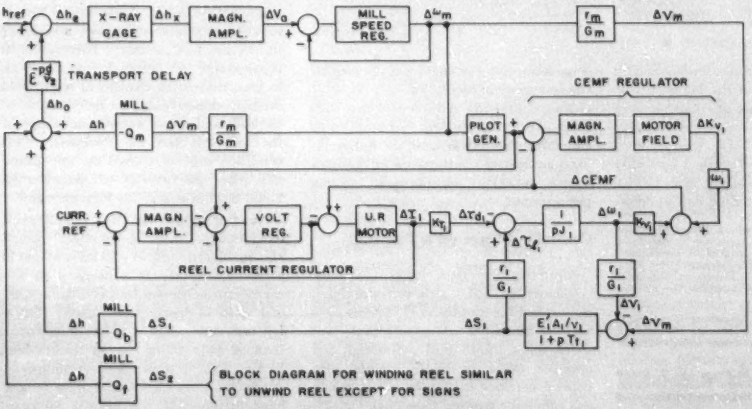


Fig. 2. System block diagram including mill stand, reels, strip, and gage control

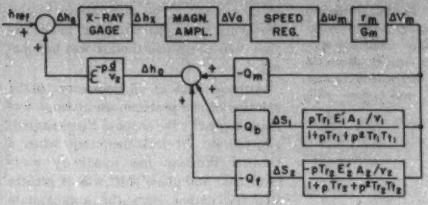


Fig. 3. Simplified block diagram of gage-control system

The three parallel paths making up $\Delta h_0/\Delta r_m$ could have been combined at this point, but the result would be a rather complex expression without much physical significance. Therefore another simplification was made by assuming that $T_1 >> T_r T_t$. This says that the feedback around the mill (see Appendix) is strong, making E' small, which was believed to be true, and will presently be shown to be true. Now the three blocks can be added, yielding

$$\frac{\Delta h_0}{\Delta r_m} = -Q_m \times \left[1+p\left\{T_{r1}\left(1+\frac{Q_b}{Q_m}\frac{E_1'A_1}{v_1}\right)+T_{r2}\left(1-\frac{Q_f}{Q_m}\frac{E_2'A_2}{v_2}\right)\right\}+p^2T_{r1}T_{r2}\left(1+\frac{Q_b}{Q_m}\frac{E_1'A_1}{v_1}-\frac{Q_f}{Q_m}\frac{E_2'A_1}{v_2}\right)\right] \Big/ (1+pT_{r1})(1+pT_{r2}) \quad (2)$$

This is still not as simple an expression as desired although its connection with reality will become apparent.

The final simplifying assumption to be made was that front tension had negligible effect on gage compared with back tension. Every operator knows that back tension is several times as effective as front tension, so this assumption is not entirely one of convenience. By making $Q_f = 0$ in equation 2 the following is obtained

$$\frac{\Delta h_0}{\Delta r_m} = -Q_m \frac{1+pT_{r1}\left(1+\frac{Q_b}{Q_m}\frac{E_1'A_1}{v_1}\right)}{1+pT_{r1}} \quad (3)$$

Since the p term in the numerator is larger than in the denominator, the mill has the characteristics of a lead circuit. That is, between the frequencies

$$1/T_{r1}\left(1+\frac{Q_b}{Q_m}\frac{E_1'A_1}{v_1}\right)$$

and $1/T_{r1}$ the gain will increase in proportion to frequency (20 decibels/decade) over what it would be without the lead effect. This effect can be explained physically as follows.

If the mill speed is varied sinusoidally at low frequencies the reel, being tied to the mill by the strip, will follow, so that unwind tension remains constant. As the frequency is increased, the reel, due to the combined effects of its inertia and the elasticity of the strip, will be unable to follow, so tension will oscillate. Since tension will tend to increase while the mill is increasing in speed, and vice versa, the net effect will be to cause a greater change in thickness than would be caused by mill speed alone. That is, the gain increases.

As frequency is further increased, the reel will eventually not change in speed at all. Above this frequency, gain will be constant again, at least insofar as the effect of the unwind reel goes. The more nearly correct equation 2 indicates that there is another down-break and another up-break, so that possibly as frequency is further increased the gain will drop down to a lower level, due to the effect of the rewind reel.

Effect of Unwind Coil Diameter

Remembering that $T_{r1} = J_1 G_1^2 v_1 / E_1' A_1 r_1^2$, equation 3 can be rewritten:

$$\frac{\Delta h_0}{\Delta r_m} = -Q_m \frac{1+p\frac{J_1G_1^2}{r_1^2}\left(\frac{v_1}{E_1'A_1}+\frac{Q_b}{Q_m}\right)}{1+p\frac{J_1G_1^2}{r_1^2}\frac{v_1}{E_1'A_1}} \quad (4)$$

Equation 4 shows that as the diameter of the coil on the unwind reel decreases, both the up-break and the down-break will decrease in frequency (because r_1^2 decreases faster than J_1), maintaining the same ratio. Thus, the frequency-response tests should be made at as nearly constant coil diameter as possible (that is, they should be completed quickly), and the coil diameter should be measured during the test.

Equation 4 also indicates that the time constants will change with speed, increasing as speed increases. This would imply that maximum speed would be the least stable condition, especially since Q_b/Q_m probably increases with speed.

Frequency-Response Tests

With this preliminary analysis completed, the next step was to run frequency-response tests. It was hoped that these would confirm the analysis and provide a number for T_{r1} and hence for E_1'.

Although several more sophisticated methods are available for determining frequency response, and these are commonly used in the laboratory, it was decided that for tests on the mill a Brush recorder would be used. This was, in part, because unlimited time would not be available since the mill would be rolling foil, and in part because considerable distortion was expected, due to inherent thickness variations and X-ray gage noise. Accordingly a Kronhite model 400-A signal

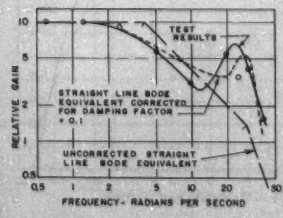

TEST RESULTS
STRAIGHT LINE BODE EQUIVALENT CORRECTED FOR DAMPING FACTOR = 0.1
UNCORRECTED STRAIGHT LINE BODE EQUIVALENT
RELATIVE GAIN
FREQUENCY - RADIANS PER SECOND

Fig. 4 (left). Gain versus frequency curves for X-ray gage

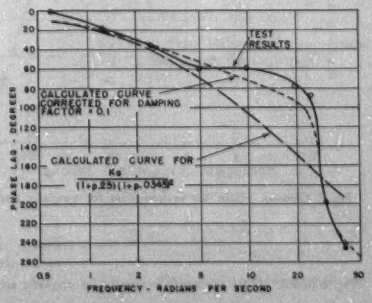

TEST RESULTS
CALCULATED CURVE CORRECTED FOR DAMPING FACTOR = 0.1
CALCULATED CURVE FOR
PHASE LAG - DEGREES
FREQUENCY - RADIANS PER SECOND

Fig. 5 (right). Phase lag versus frequency curves for X-ray gage

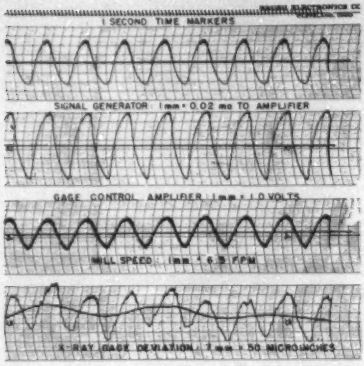

1 SECOND TIME MARKERS
SIGNAL GENERATOR 1 mm = 0.02 mo TO AMPLIFIER
GAGE CONTROL AMPLIFIER 1 mm = 1.0 VOLTS
MILL SPEED 1 mm = 6.5 FPM
X-RAY GAGE DEVIATION 1 mm = 50 MICROINCHES

Fig. 6 Typical recording taken during frequency-response test. The approximate zero line is drawn in to aid in determining average amplitude and phase shift

generator and a 6-channel Brush oscillograph were used for the tests.

Mill Description

The mill on which the frequency-response tests were run was a 10-inch and 24- by 48-inch single-stand nonreversing foil mill. The mill stand is driven by a 1,000-hp (horsepower) 575/1,150-rpm d-c motor through a pinion stand. The unwind reel is powered by two 50-hp 200/800-rpm motors, and the rewind reel is driven by two 50-hp 400/1,600-rpm machines, both through reduction gears. Each reel and the mill stand has its own generator.

X-Ray Gage Response

The first step was to measure the frequency response of the X-ray gage, since it is most important to know this to determine the response of the mill. It was discovered after the tests were run on the mill that the amplifier used with the recorder was disturbing the operation of the gage. Although this did not invalidate the mill tests it did not give accurate information on the gage, so the interaction was eliminated by using an isolating transformer and another test run on the gage.

Fig. 4 shows the resulting gain curve, and Fig. 5 the phase lag. From previous tests on the same design gage it was known normally to have a resonant peak at about 25 rad/sec (radians per second). Hence, additional data were taken near this frequency, although the other points were all approximately an octave apart. It will be noted that all points seem to fall on a plausible curve except the one at 25 rad/sec. The gage output at this frequency was quite irregular, indicating that something unorthodox was happening.

Although not at all necessary to the solution of the problem, an attempt was made to justify the shape of the measured-gain curve. It is known that when a transfer function has imaginary roots both gain and phase shift will be greater than calculated from the approximate transfer function (or measured with a Bode ruler from the straight-line approximation, as was done here).[3] Using the curves given in reference 2 the corrected gain- and phase-lag curves shown in Figs. 4 and 5 were obtained, using a damping factor of 0.1.

While the correlation with test data is still not very good this at least makes the data seem more plausible. In connection with correlation it should be recognized that because of the size of the recorder charts used and width of traces, amplitudes could be read only to an accuracy of about 5 or 6% and phase angles to about 9 degrees at best. Added to these errors are those due to nonlinearities of system components.

Tests on Mill

After getting the first-mentioned gage-response data the signal generator and recorder were set up for the big test. Referring to Fig. 3, the circuit was opened between the X-ray gage and the magnetic amplifier and the signal generator was connected to the input of the amplifier. The quantities recorded were, in addition to the signal generator output, ΔV_a, Δv_m, and Δk_r. The lowest frequency used

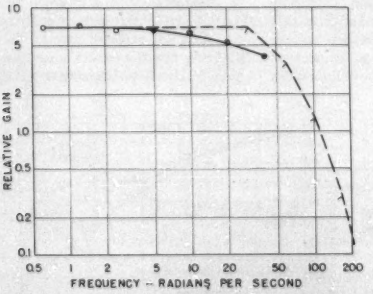


Fig. 7 (above). Gain versus frequency for magnetic amplifier

O = test points
- - - = straight-line Bode approximation

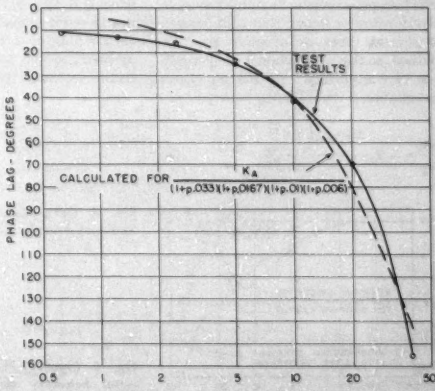


Fig. 8 (right). Phase lag versus frequency for magnetic amplifier

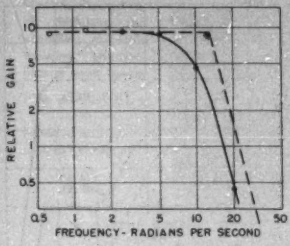


Fig. 9. Gain versus frequency for mill motor and control

O = test points
- - - = straight-line Bode approximation

was 0.1 cps (cycle per second), as any lower frequencies would take too much time, and it was desirable to cover a complete range of frequencies with as nearly as possible the same coil diameter. Above 1 or 2 cps the attenuation increased very rapidly, putting an upper limit on the frequency range. Fig. 6 shows a typical recording for one frequency, and gives some idea of the difficulty of obtaining precise data from the gage. At lower frequencies the wandering and gage noise seriously distort the supposedly sinusoidal output, and at higher frequencies the amplitude becomes so small as to make accurate measurements impossible. Nevertheless, one of the purposes of this paper is to show that even under such circumstances it can be feasible to secure intelligible results.

From the charts thus obtained amplitude and phase shift versus frequency were plotted for $\Delta V_a/\Delta V_s$, $\Delta v_m/\Delta V_a$, and $\Delta h_z/\Delta v_m$. From the first two, approximate transfer functions were obtained by cut-and-try methods. That is, based on the gain curve a straight-line Bode approximation was drawn. Its phase shift was then measured with a Bode ruler and plotted on the phase-shift curve. By successive trials a compromise was reached that represented the best fit of both curves. The results are shown in Figs. 7–10.

Alternate Methods

These transfer functions could have been obtained by calculation accurately enough for designing stabilizing circuits, although the frequency-response test method was used since the test was being run anyway. However, as the analysis showed, obtaining the transfer function of the mill by calculation would be very difficult if not impossible with the knowledge available.

It should also be pointed out that it is not absolutely necessary to measure phase shift to obtain usable results. That is, a transfer function accurate enough for designing stabilizing circuits can often be deduced from the gain versus frequency curves. This is a great timesaver, as determining phase shift from oscillograph charts is a tedious job.

Mill Transfer Function

To obtain the mill transfer function, $\Delta h_z/\Delta v_m$ was first plotted. However, what was required was $\Delta h_0/v_m$. This was obtained by dividing $\Delta h_z/\Delta v_m$ by the X-ray gage gain, $\Delta h_z/\Delta h_0$, previously determined. Fig. 11 shows the results.

To obtain phase shift it was necessary to subtract the phase shift contributed by the gage, by the noise filter used for recording purposes, and by the transport time between roll bite and gage. Since this required the subtraction of two large numbers, individual points were not subtracted, but smooth curves were drawn through each set of data. These are shown in Fig. 12. Fig. 13 was then drawn by subtracting the two curves. Again, the best fitting straight line approximation was determined by trial, and it and its phase shift are plotted on Figs. 11 and 13 respectively.

It is interesting to note that while the measured gain has the shape predicted by equation 3 over the frequency range it was possible to use, the best fit for the phase-shift curve was a transfer function more nearly resembling equation 2. However, in the frequency range of importance to the regulator the simpler expression is probably adequate, and was used. This made it possible to take into account the change in coil diameter.

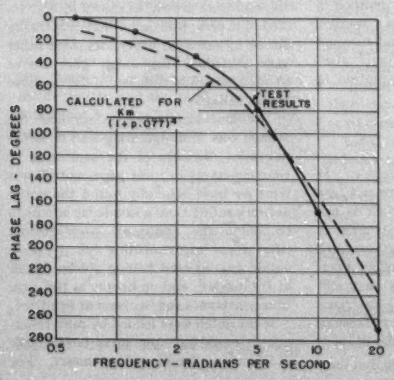


Fig. 10. Phase lag versus frequency for mill motor and control

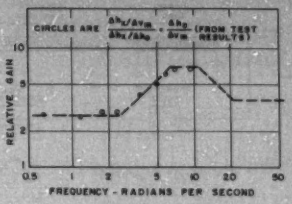


Fig. 11. Gain versus frequency for mill
- - - = straight-line Bode approximation

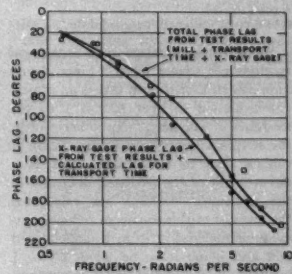


Fig. 12. Phase lag versus frequency from tests on mill and X-ray gage

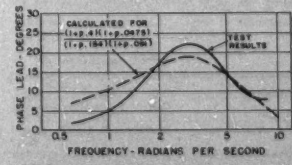


Fig. 13. Phase lead versus frequency for mill

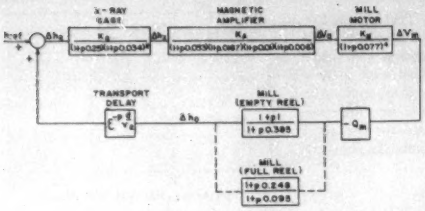


Fig. 14 (left). Final block diagram of system, not including stabilizing circuits

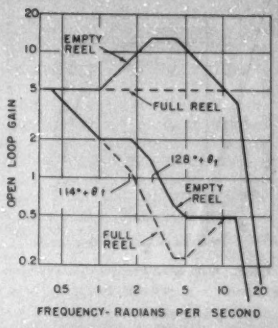


Fig. 15. Bode diagrams for the gage control. Upper curves are without stabilizing circuits. Note that at full reel the up-break of the mill is cancelled by the gage time constant. Lower curves include effect of notch network and rate feedback

As has been pointed out, the mill transfer function varies with coil diameter, whereas the tests were made at approximately a constant diameter. Thus it is necessary to calculate the time constants at full and empty unwind reel. The results are given in Table I.

Discussion of Reel Time Constant, T_{r1}

As shown in Fig. 11, the time constant T_{r1} was seen to be about 0.154 second. Since all numbers in the expression for T_{r1} are known except E_1', we can use this result of the frequency-response tests to determine a value of E_1', the "apparent" modulus of elasticity.

Thus, $E_1' = J_1 G_1^2 v_1 / T_{r1} A_1 r_1^2$. For the conditious under which the frequency-response tests were taken the following numbers apply:

$J_1 = 14.5$ slug ft (feet)^2
$T_{r1} = 0.154$ sec
$G_1^2 = 7.76$
$A_1 = 0.1$ in. (inch)^2
$v_1 = 7.7$ ft/sec
$r_1^2 = 0.69$ ft^2 (average measured during test)

From these calculate $E_1' = 81{,}000$ psi (pounds per square inch). Since aluminum has an E of 10^7 psi, this is an apparent reduction of over 120 to 1 in the modulus of elasticity. In other words, the strip acts more as if it were made of rubber than of aluminum. This is a direct result of the feedback discussed in the Appendix. Phillips has calculated the amount of this apparent reduction in modulus for steel strip in a tandem cold mill and found it to be about 35 to 1 between the last stand and the reel, and between 240 to 1 and 335 to 1 between stands.[3]

Obviously, it is because E_1' is so low that the reels affect the gage control at frequencies in the vicinity of crossover. There are other consequences of what might be called this basic fact of rolling. For example, it is probably what makes a tandem mill possible, for without it, it would be difficult to keep interstand tensions within reasonable bounds.

It is partly because E_1' is so low, and hence T_{r1} is so long, that controlling gage on a foil mill by adjusting back tension does not appear to be as satisfactory as using mill speed. It can be shown that, when back tension is used, the block diagram will be very similar to that existing when mill speed is used except that the inertia time constant of the mill motor is replaced by T_{r1}, the time constant between reel torque and back tension. If the latter is larger, as will frequently be the case, the system response will be poorer. In addition, the lead effect of the unwind reel, which frequently improves response, will not be present.

The time constant T_r, and the phenomenon of reduced apparent modulus of elasticity, is also an important factor in tension regulators. It can adversely affect both stability and gain if not understood and taken into account.

Measuring E′

The use of frequency-response tests is not the easiest or most direct method of measuring E′. A more direct method is to make tension changes and measure the resulting change in v_r. If mill speed is held constant, then from equation 18 in the Appendix,

$$E_1' = \frac{-\Delta S_1}{A_1} \frac{v_1}{\Delta v_1} \qquad (5)$$

for the unwind reel. One might expect that Δv_1 would be too small to measure with simple equipment, but such is not the case. On several rough tests on both aluminum and steel mills the authors have measured reel speed changes of as much as 20% with tension changes that still permitted rolling. Although no very accurate tests have been made, the results have at least confirmed the order of magnitude expected for E_1'.

Table I

Diameter	38 in. (full)	20 in. (test)	12 in. (empty)
J_1/r_1^2	13	21	52.5
Up-break	0.248	0.4	1.0 sec
Down-break, T_{r1}	0.005	0.154	0.385 sec

Gage-Control Performance

Stabilizing the Regulator

In addition to increasing our understanding of the system, the real reason for this work has been to improve the performance of the gage control and learn how to design such controls for best performance.

The first step in stabilizing the system was to draw the final block diagram, using the transfer functions determined from the frequency-response tests. This is shown in Fig. 14. The d-c gain figures are not shown, since the individual gains are not important. In practice, the amplifier gain is made sufficiently large to give the desired open-loop gain.

From the block diagram, the straight-line Bode diagrams for both empty and full reel were drawn, as shown in Fig. 15. The d-c gain was somewhat arbitrarily set at 5, based on experience with other gage controls. Any loop gain above about 3 is usually sufficient if the response is fast enough.

The first step taken to stabilize the system was to design a lag-lead or notch network. Because of the multitude of time constants in most gage-control systems we have generally found the notch network rather than a simple lag circuit to be worth the additional complication. After some experimenting (on paper) a notch was selected having a down-break at 0.4 rad/sec, and up-breaks at the gage time constant (4 rad/sec) and at 5 rad/sec.

If the notch were loaded by pure resistance it would have one more down-break at 75 rad/sec for the values chosen. For

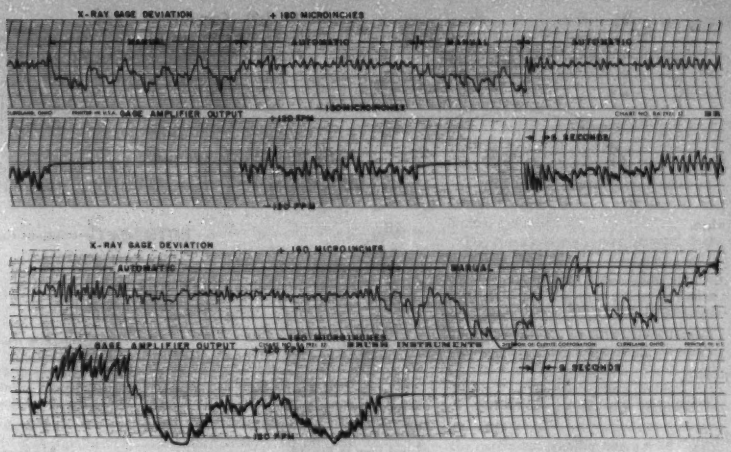


Fig. 16. Recordings of thickness and amplifier output showing effectiveness of gage control

Nominal thickness = 0.002 in.
Nominal mill speed = 450 ft/minute

circuit reasons, it was necessary to put the notch network in series with a relatively low resistance and a relatively high inductance of a magnetic-amplifier control winding. This causes an additional down-break, so the final notch has down-breaks at 30 and at 62.5 rad/sec. The extra down-break is helpful in this instance in attenuating the 120-cycle ripple from the first stage magnetic amplifier.

A rate feedback, producing a lag, was added around one stage of magnetic amplifier partially to cancel the lead caused by the unwind reel, although the Bode diagram shows that stability can be achieved without this on this mill. The resulting stabilized Bode diagram is also shown on Fig. 15.

Similar frequency-response tests were made on another mill, which had a higher reel inertia and a higher reel-gear ratio. At least partially because of this the time constant T_{r1}, was considerably longer, the change from full to empty reel was much greater, and the total gain increase was about 6 to 1. The gage control was installed first on this mill, and it is the one in which the rising gain was suspected, prompting this study. On it, both the Bode diagram and field experience confirm that the rate feedback is necessary to obtain stability with reasonably fast response. Incidentally, the same notch network is used for this and a number of other mills, leaving only the rate feedback and over-all gain to be adjusted in this field.

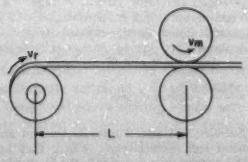


Fig. 17. Unwind reel and mill stand with no reduction

Performance

None of the gage controls have been operating for enough time to have accumulated very complete data on the over-all improvement contributed by the gage control. Fig. 16 shows part of two recordings made at the time the stabilizing circuits were being adjusted. This is the mill on which the frequency-response tests described here were made. The upper chart seemed to represent average quality foil of this thickness, while the lower chart is an example of unusually poor foil. In both instances, it will be noted that the gage control makes an appreciable improvement over manual control.

Later checks, while not yet conclusive, indicate that the automatic control limits gage variation to approximately one-half that experienced when using manual control.

Side benefits, which have occurred, are fewer breaks in the strip and consequently increased mill production.

Conclusions

On the basis of the work done on two foil-mill gage controls it is concluded that frequency-response tests on a complete system can be a very useful tool to supplement frequency-response analysis.

In summary, the tests described here accomplished at least the following three objectives:

1. They confirmed the simplified analysis of the system, thus adding to the understanding of the gage control.

2. They provided experimental verification for the theoretically predicted apparent reduction in modulus of elasticity thus increasing knowledge and understanding of all metal-rolling processes.

3. They made it possible to design new stabilizing circuits for the foil-mill gage control and thus improve its performance appreciably.

Appendix

Derivation of Tension Transfer Function

The following is not a rigorous proof. It is a simplified explanation intended primarily to provide a clear physical picture of relationships that actually exist.

Consider first an unwind reel feeding a mill stand, which for the moment is not taking any reduction. This is illustrated in Fig. 17.

By definition

$$E = \frac{\text{Stress}}{\text{Strain}} = \frac{S/A}{e/l}$$

from which

$$e = \frac{Sl}{AE} \qquad (6)$$

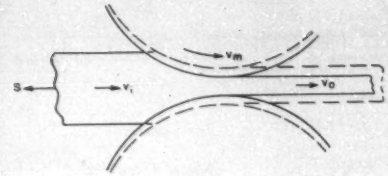


Fig. 18 (left). Cross section of mill-roll gap showing two different reductions

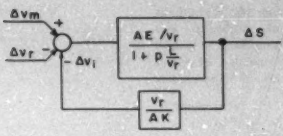


Fig. 19 (right). Block diagram of mill

where

E = modulus of elasticity, lb (pounds)/in.^{2}
S = tension, lb
A = strip cross section, in.^{2}
e = elongation, ft
l = unstressed length, ft

From Fig. 17, L is the stressed length. From the above definitions we can write

$$l+e=L \tag{7}$$

Substituting equation 6 into equation 7

$$l+\frac{Sl}{AE}=L$$

and

$$l=\frac{L}{1+S/AE} \tag{8}$$

Differentiating both sides of equation 8

$$dl=\frac{-L/AE}{(1+S/AE)^2}\,dS \tag{9}$$

Typically, $S/A=5{,}000$ lb/in.^{2} and $E=10^7$ lb/in.^{2}, so that $(1+S/AE)^2=(1+0.0005)^2=1+0.00000025$.

Thus little error is introduced by neglecting the term in parentheses and writing

$$dl=\frac{-L}{AE}\,dS \tag{10}$$

When the mill is rolling, strip, which for this analysis can be considered to be unstressed, is continually being unwound from the reel and stressed. The equivalent length of stressed strip unwound in time dt is

$$dl_i=v_r dt(1+S/AE) \tag{11}$$

Also, strip is continually leaving the mill after being stressed and elongated. The length leaving the mill in time dt is

$$dl_0=v_m dt \tag{12}$$

The change in length $dl=dl_i-dl_0$. Subtracting equation 12 from equation 11 yields

$$dl=dl_i-dl_0=(v_r+v_r S/AE-v_m)dt \tag{13}$$

Substituting equation 10 in equation 13

$$-\frac{L}{AE}\,dS=(v_r+v_r S/AE-v_m)dt \tag{14}$$

Substituting $p=d/dt$ produces

$$-Lp S/AE=v_r+v_r S/AE-v_m \tag{15}$$

Rearranging equation 15 and solving for S yields

$$S=\frac{\dfrac{AE}{v_r}(v_m-v_r)}{1+pL/v_r} \tag{16}$$

For small changes we can write

$$\Delta S=\frac{\dfrac{AE}{v_r}(\Delta v_m-\Delta v_r)}{1+pL/v_r} \tag{17}$$

Equation 17 is true if the mill is taking no reduction. Now imagine that the rolls are brought closer together, so that some reduction is taken, as represented by the dashed lines in Fig. 18.

For this explanation, the neutral point, or point of no slip, will be assumed to be at the exit of the roll gap, so that $v_0=v_m$. In actuality the neutral point can be anywhere in the roll gap, depending on reduction, tensions and other factors, but this does not change the result qualitatively.

Now, let us start with a well known fact that requires no proof; i.e., that if back tension is increased the reduction will increase. The condition with increased back tension is represented in Fig. 18 by solid lines.

By inspection of Fig. 17, as well as by equation 17, to increase tension v_r must be reduced. As soon as this is done, and the reduction increases, v_i must also become smaller, if $v_0=v_m=$constant. This partially relieves the tension that was established. To re-establish it v_r must be reduced further. Thus, in the practical case where tension affects reduction the tension obtained is no longer correctly determined by equation 17. The change in v_i is in effect a negative feedback from ΔS to $\Sigma\Delta v$, which can be represented in block diagram form as shown in Fig. 19.

Without the feedback this is a block diagram representation of equation 17. In order to give the feedback the correct dimensions, so that the loop gain is dimensionless, we can make the feedback block v_r/AK. The feedback factor K, which Phillips has called the "Modulus of Plasticity,"[4] has the same units as E. Closing this loop results in

$$\frac{\Delta S}{\Delta v_m-\Delta v_r}=\frac{\dfrac{\dfrac{AE}{v_r}}{1+p\dfrac{L}{v_r}}}{1+\dfrac{\dfrac{E}{K}}{1+p\dfrac{L}{v_r}}}=\frac{\dfrac{\dfrac{AE}{v_r}}{1+\dfrac{E}{K}}}{1+p\dfrac{\dfrac{L}{v_r}}{1+\dfrac{E}{K}}}=\frac{\dfrac{AE'}{v_r}}{1+pT_t} \tag{18}$$

where

E' = "apparent" modulus of elasticity

$$=\frac{E}{1+\dfrac{E}{K}}$$

and T_t = transport time constant

$$=\frac{\dfrac{L}{v_r}}{1+\dfrac{E}{K}}=\frac{L}{v_r}\frac{E'}{E}$$

Equation 18 is represented in block diagram form in Fig. 2. It has been derived for the unwind side of the mill, but by the same method a similar expression can be derived for the rewind side.

Nomenclature

A = strip cross-sectional area, in.^{2}
d = distance from mill to x-ray gage, ft
E' = "apparent" modulus of elasticity, lb/in.^{2}; see Appendix
G = gear ratio between reel motor and reel
G_m = gear ratio between mill motor and mill
h_{ref} = X-ray gage thickness reference
h_0 = strip thickness leaving mill
h_e = thickness error = $h_{ref}-h_0$
h_x = X-ray gage output (deviation)
I = reel motor current
J = total reel inertia referred to motor, slug ft^2
K_A = magnetic amplifier-gain factor
K_G = X-ray gage-gain factor
K_M = mill-motor gain factor
K_V = reel-motor voltage constant, volt/revolution/sec
K_T = reel-motor torque constant, lb-ft/ampere
L = reel-to-mill distance, ft
p = heaviside operator = d/dt
$Q_m=\Delta h_0/\Delta v_m$
$Q_0=\Delta h_0/\Delta S_1$
$Q_f=\Delta h_0/\Delta S_2$
r = radius of reel (including coil), ft
r_m = radius of driven mill roll, ft
S = tension, lbs
T_r = tension time constant = $JG^2\omega/E'Ar^2$
T_t = transport time constant; see Appendix
τ_d = developed motor torque, lb-ft
τ_t = tension torque, lb-ft
v, v_r = peripheral speed of reel, ft/sec
v_m = mill roll peripheral speed, ft/sec
V_a = magnetic amplifier output voltage
V_s = signal generator, output voltage
ω = reel-motor speed, revolutions/sec
ω_m = mill motor speed, revolutions/sec
θ_t = phase shift due to transport time between mill and X-ray gage

$$=\frac{180^\circ}{\pi}\frac{d}{v_r}\text{ per rad/sec}$$

Subscripts 1 and 2 refer to unwind and rewind reel respectively. Where they are omitted from reel quantities the expression can be applied to either reel by using the appropriate subscript.

References

1. Rolling of Thin Strip, M. D. Stone. *Proceedings*, Association of Iron and Steel Engineers, Pittsburgh, Pa., 1953, pp. 116–28.

2. Servomechanisms and Regulating System Design, Vol. I (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., New York, N. Y., 1951, pp. 310–14.

3. Analysis of Tandem Cold Reduction Mill with Automatic Gage Control, R. A. Phillips. *AIEE Transactions*, vol. 73, pt. II, 1956 (Jan. 1957 section), pp. 355–63.

4. Process Analysis Plus Analog Simulation Yields Better Mill Controls, R. A. Phillips. *Control Engineering*, New York, N. Y., May 1958, pp. 113–18.

The Optimum Control of Multiactuator Systems

IAN McCAUSLAND
ASSOCIATE MEMBER AIEE

MUCH of the theory of control systems deals with the problem of reducing to the lowest possible value the dynamic error in single-variable control systems. In the study of that problem it is generally assumed that any conceivable method of compensation is justified, provided that it allows the error, as measured in accordance with some criterion, to be reduced; in other words, the purpose of the controller is to reduce the error at all costs.

There are, however, many systems in which other considerations may become more important than reduction of error. Such a system might be a process involving the use of large amounts of power, in which considerable effort must be exerted in order to achieve a change in the controlled or output variable. In such a process there may often occur a system state such that the benefit derived from a small change in the output variable requires more effort than it is worth. In a case like this it is better to be satisfied with the existing system state than to attempt to reduce the error between the actual value of the output variable and the desired value.

Many industrial processes which require the use of effort to obtain a desired result are characterized also by the fact that the output variable can be altered by the adjustment of any of several input quantities. It is as if there were more actuators than one, and it is clear that in a process of this type a desired change in the output can be accomplished in many different ways, depending on how the total change in output is shared among the separate actuators. In general some of these methods of sharing will be better than others, so that the control problem is the problem of choosing a suitable division of the load. It is obvious that there is no problem of choice if all the actuators which are available are identical, for in such a system the correct decision would be to treat all the actuators alike. If the change in the output variable can be achieved without effort, by any one actuator, there is no problem of choice because the actuator which can accomplish the result without effort can be used exclusively.

The criteria by which the best division of effort is chosen may involve many considerations. In some cases it may be a question of sharing a given power load in the best way; in others it may be a question of minimizing wear and tear on equipment, while in still other cases the effects of changes in the input quantities on certain secondary output variables may be significant. The common feature about all such processes is that it is not sufficient to concentrate on keeping only one variable at its desired value, but it is necessary also to recognize the importance of other variables. That is, instead of seeking a particular value of a particular variable, the controller should seek the most desirable over-all state for the system under its control.

A basic theory of control of multiactuator systems is described in this paper, with a brief outline of a portion of economic theory which deals with a similar type of problem. Methods of automatically realizing the optimum system state are described, and some brief suggestions concerning practical applications are included.

Division of Effort Among Actuators

The problem of control in a multiactuator system is the problem of using actuators of limited capabilities in the best over-all way. This problem is remarkably similar to one of the fundamental problems of economics, i.e., the administration of scarce resources in such a way as to achieve the best possible result. In economics, as in the control system, the problem becomes meaningless if the resources are available in unlimited quantity.

The theory of consumer demand deals with the problem in economics which corresponds to the multiactuator control problem. In this theory it is assumed that a consumer derives a certain amount of utility from the goods he purchases, the amount of utility being a function of the quantities of the goods acquired. The consumer's problem is to spend his income on the commodities available, in such proportions as to derive the maximum possible amount of utility. The theory involved in this maximization will be outlined shortly; before doing this some ideas on the optimum state of a multiactuator system will be considered. The most important idea is that of a "penalty function," which will now be described.

Penalty Function

Just as in economics the consumer wishes to maximize the utility, so in the control system it is desired to use the available actuators in such a way as to maximize or minimize some quantity which represents the performance of the system. Perfect control in the single-variable system is represented by the condition that the error is zero at all times, and the quality of performance is assessed by some function of the error, which represents the amount by which the system falls short of the ideal. Similarly, in the multiactuator system, perfect control is represented by the condition that all inputs and outputs remain at their desired values at all times. Therefore, instead of trying to maximize the utility, it is more convenient to think in terms of minimizing the amount by which

Paper 59-198, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 3, 1958; made available for printing December 5, 1958.

Ian McCausland is with the University of Toronto, Toronto, Ontario, Canada.

The author gratefully acknowledges the advice and encouragement of Prof. J. M. Ham of the University of Toronto, under whose guidance this investigation was carried out. The author's gratitude is also extended to the Canadian Westinghouse Company for financial assistance during the period in which the work was performed.

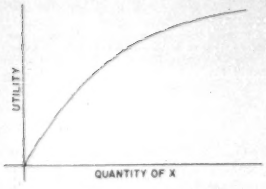


Fig. 1. Utility of commodity x

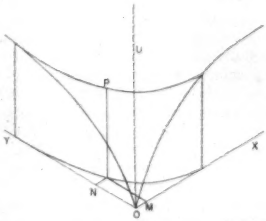


Fig. 2. Utility surface for two commodities x and y

the system falls short of the ideal. The magnitude of this shortcoming is a negative "score" in the system's game against nature, so it is proposed that this quantity be denoted by the term "penalty." That is to say, the purpose of the controller is to minimize the operating penalty. The penalty will usually be a function of the instantaneous system state, and two similar controllers would be compared by comparing the average penalty for the two cases, in the presence of input signals and disturbances of particular types. The nature of the penalty function will be discussed later.

Theory of Consumer Demand

As has already been pointed out, the controller which has several actuators available to obey its commands is in a somewhat similar position to the consumer who has several commodities available to supply his wants. Therefore it is possible to learn something about the situation by studying the consumer who spends his income in such a way as to maximize the utility which he derives from these commodities. This subject has been thoroughly studied,[1] and part of the theory is outlined here because it forms a useful basis for the study of multiactuator systems.

The amount of utility which is derived from one commodity x can be plotted as a function of the quantity of the commodity, as shown in Fig. 1. The shape of the curve is explained by the fact that the incremental utility generally decreases as the quantity of the commodity is increased. In the same way, when dealing with two commodities x and y, a utility surface can be drawn, as shown in Fig. 2; the utility is measured on the vertical axis and is denoted by the symbol u. The vertical height of any point P on the surface therefore represents the amount of utility obtained by the contributions of amounts OM of x and ON of y.

Indifference Curves

The contour lines of the utility surface, when projected onto the xy plane, form a family of curves as shown in Fig. 3. These curves are called "indifference curves." The magnitude of the slope of the indifference curve at any point is the amount of y which is needed by the consumer to compensate him for the loss of a small unit of x. The amount of utility lost, due to the loss of a small unit of x, is the amount of x lost multiplied by the incremental utility of x. A similar relationship holds for y, and these lead to the following equation:

$$\frac{dy}{dx} = -\frac{\partial u}{\partial x} \Big/ \frac{\partial u}{\partial y} \qquad (1)$$

That is, the magnitude of the slope is equal to the incremental utility of x divided by the incremental utility of y. As one moves along the indifference curve in the direction of increasing x (and hence of decreasing y) $\partial u/\partial x$ decreases and $\partial u/\partial y$ increases, so that the magnitude of the slope decreases, i.e., the indifference curve is everywhere convex to the origin.

The condition of maximum utility for a given income may be found from the indifference map. This is illustrated by the graph in Fig. 4. On the diagram, OL is the quantity of x which would be purchased if the whole income were spent on x, and similarly OM is the corresponding quantity of y. Any point on the straight line LM therefore represents a quantity of x and a quantity of y which together would cost the whole income. Starting at the point P, if one moves along the straight line LM in the direction of increasing x, one leaves the indifference curve shown and moves to one of higher utility. Hence P cannot represent the condition of maximum utility. However, starting at the point Q, at which the straight line is tangential to an indifference curve, a movement along the line in either direction causes the utility to decrease; hence Q represents the condition of maximum utility for that particular income. It is easy to show that, at the condition

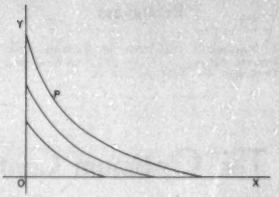


Fig. 3. Indifference curves for x and y

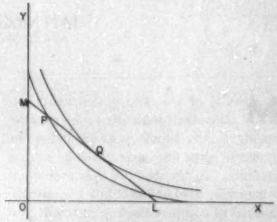


Fig. 4. Indifference curves, illustrating the condition of maximum utility

of maximum utility, the incremental utilities of the two commodities are directly proportional to their prices.

It is interesting to note that it is not necessary to know the actual value of utility which corresponds to a particular indifference curve; it is sufficient to know which direction represents an increase in utility and which represents a decrease. That is, it is only necessary to have a scale of preference, rather than a quantitative knowledge of the utility.

Application to Control Systems

The 2-actuator control problem is similar to the problem of the consumer to whom two commodities are available. The basic 2-actuator system is shown in Fig. 5; it is desired to control the output variable c by one or both of the actuators G_1 and G_2. In general the variables inside the dotted rectangle are unobservable, i.e., it is not possible by observation of the output side to determine how much of c has been contributed by which actuator. However, each transfer function can be determined directly by varying one input and keeping the other fixed at zero, so that it is possible to observe the inputs under steady-state conditions and thereby, if the system is linear, to know the contributions made by the respective actuators to the output quantity.

In this preliminary exposition of the theory of control of multiactuator systems it is assumed that all reference inputs

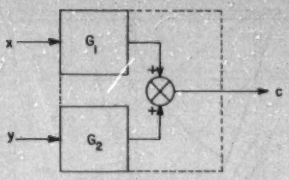


Fig. 5. The basic 2-actuator control system

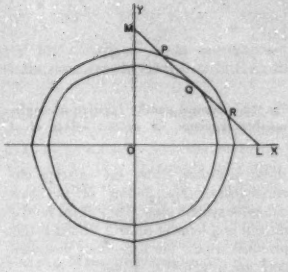


Fig. 6. Indifference map for 2-actuator system

and disturbance inputs to the system are of very low frequency, so that it is possible to consider the function of the controller as being to alter the steady-state operating point of the system to the condition of minimum penalty whenever it departs from that condition. Every transfer function which appears in an equation in this paper will therefore be taken to mean the value of the transfer function at zero frequency. Thus the output c of the system shown in Fig. 5 is given by

$$c = G_1 x + G_2 y \qquad (2)$$

In the case of a control system a set of indifference curves of penalty can be drawn, which are similar to the indifference curves of utility in the economic example except that the penalty curves are normally concave to the origin. The curves are concave to the origin because in most practical cases large errors are relatively more costly than small errors, i.e., the penalty contributed by a particular variable usually increases more rapidly as the value of the variable departs further from the desired value. Another difference between the indifference curves of the economic system and those of the control system is that those for the control system can be drawn in all four quadrants of the xy plane, instead of in the first quadrant only as in the economic example.

Fig. 6 shows a hypothetical indifference map for a 2-actuator control system. Here, similarly to the economic case, OL represents the amount of input x which would be required to give output c if y were set at zero; similarly OM is the corresponding quantity for y, i.e., $OL = c/G_1$ and $OM = c/G_2$. Any point on the line LM therefore represents a pair of inputs x and y which if acting together would give an output c. As before, it may be seen that the point P in Fig. 6 is not a point of minimum penalty, but that Q represents the desired condition. It is required to design a controller which will adjust the inputs x and y in such a way that the point of minimum penalty is reached.

That the control system should be able to seek the point of minimum penalty, it is necessary that information about the relationship between the penalty and the system variables should be available to the controller. Therefore an essential feature of such a controller is a function generator whose inputs are the significant variables of the system and whose output is the penalty. Such a function generator is shown in Fig. 7. It is assumed that the state of the system is such that the output is at the desired value c, but that the inputs x and y require adjustment so as to minimize the penalty. The purpose of the function generator is the following: the inputs to the function generator are altered by an auxiliary signal source, as shown in Fig. 7, in such a way that the apparent operating point, as seen by the function generator, moves alternately to one side and to the other of the actual operating point. This can be achieved by adding small sinusoidal signals $aG_2 \sin \omega t$ and $-aG_1 \sin \omega t$ to the signals x and y respectively, and using the two sums as

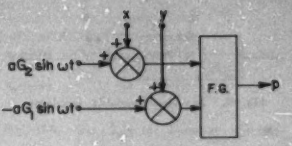


Fig. 7. The penalty function generator

the inputs to the function generator as shown in Fig. 7. The path of the apparent operating point is along the line LM in Fig. 6, as shown by the following equation:

$$(x + aG_2 \sin \omega t)G_1 + (y - aG_1 \sin \omega t)G_2 = x.G_1 + y.G_2 = c \qquad (3)$$

Therefore, with the sinusoidal inputs shown, the apparent operating point oscillates about the actual operating point, along the line LM, with simple harmonic motion of small amplitude. The output of the function generator therefore varies in a way which depends on the position of the actual operating point. For example, at point P (see Fig. 6) when the apparent operating point is moving in the direction of increasing x, the penalty is decreasing; therefore the incremental output of the function generator is opposite in phase to the incremental x input. The total output of the function generator is a steady value plus an alternating component. At the point R the incremental output is in phase with the incremental x input, while at Q the incremental output is negligible.

If the output of the system is to be kept at a fixed value c, there is only one degree of freedom for the inputs, represented by the line LM in Fig. 6. Therefore the relative adjustment of the inputs can be made by a single-variable actuator. If a voltage supply $b.\sin(\omega t - \pi/2)$ is used to energize the reference phase of a 2-phase

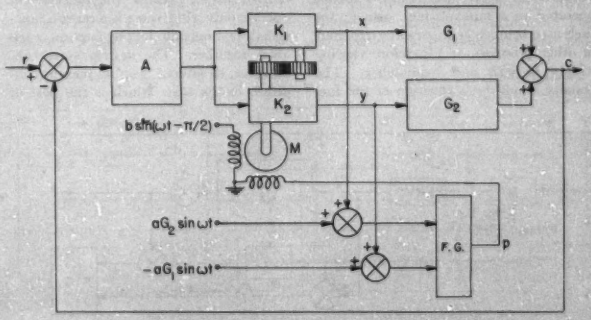


Fig. 8. Block diagram of complete 2-actuator servomechanism

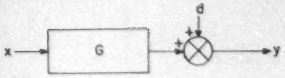


Fig. 9. The single-variable regulator

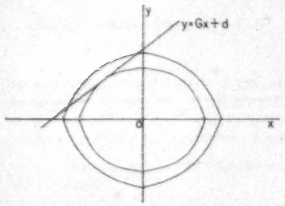


Fig. 10. Indifference map for single-variable regulator

servomotor, and if the control phase is energized by a voltage proportional to the incremental output of the function generator, the motor will run in one direction when the operating point is on one side of the minimum-penalty point, and in the opposite direction when it is on the other side. At the point of minimum penalty the motor will develop no torque. A block diagram showing a complete servomechanism operating on this principle is shown in Fig. 8. The high-gain amplifier A ensures a low steady-state error, while the servomotor M adjusts the two series gains K_1 and K_2 in opposite directions to each other in order to provide the inputs x and y in the correct proportions for minimum penalty.

It is conceivable that a control system of this type might be used to control the temperature of a furnace, the controlled variable being temperature, and the inputs being the amounts of heat supplied by two fuels. Open-hearth and soaking-pit furnaces in the steel industry can be operated in a multiple-fuel fashion, the fuels used being in many cases by-products of other processes and therefore varying widely in cost and availability. The penalty would be a function of the fuel costs, and of the burner efficiencies which would vary because of the large variations in the rates of fuel flow. It is possible that certain problems of similar type might be amenable to linear programming methods of solution, but the variation of burner efficiencies with rates of fuel flow makes such methods unsuitable in this case because of the nonlinear relationships involved.

Single-Variable Regulator

In certain physical processes the single-variable regulator problem may be similar to the 2-actuator problem. For example, in processes where the input is at a high power level, the cost of altering the input in order to neutralize a small disturbance in the output variable may be comparable to the cost of allowing the disturbance to remain uncorrected. One way of controlling such a process would be to use an on-off controller with a dead zone of suitable size. An on-off controller has the effect of separating all disturbances into two categories, the significant and the insignificant, and it controls the system in such a way that a disturbance whose magnitude is many times the dead zone is corrected to the same value as one which is just greater than the dead zone, even though the demands on the costly input may be vastly different in the two cases. On the other hand, a human operator of such a system, if he had a thorough knowledge of all the factors involved, could use his knowledge and judgment to decide at what point a small improvement was not worth the extra effort involved.

In other words, it is not always sufficient to control the output of a system to its best possible value; the input must also be considered. It is not possible to adjust every variable in the system to the desired value, because the process will operate only when there is a certain functional relationship linking certain variables together. The system state can, however, be altered from its present condition to the state which is the best of those available under the existing circumstances. The finding of the best available system state is one of the fundamental problems of control, and it is this problem which the regulators described here are intended to solve.

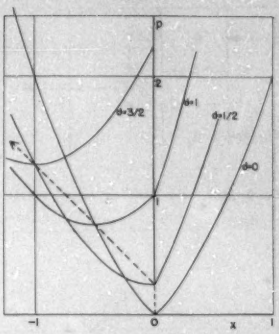


Fig. 12. Typical penalty function of single-variable regulator for various values of d

Consider the single-variable regulator shown in Fig. 9. The variables x and y are taken to represent the incremental values of the input and output respectively; i.e., the nominal system state is $x=0$, $y=0$. It is desired to control the system state in the presence of disturbances represented by d.

Fig. 10 shows the indifference map for the system, i.e., the contour lines of the penalty function $p(x,y)$. For a particular disturbance d, the available system states are represented by the straight line $y=G.x+d$. Therefore the point of minimum penalty is found as before.. An automatic method of adjusting the system to the optimum operating point is shown in Fig. 11. The block G' is a low-power steady-state model of the block G, in the sense that the transfer function of G' is at all frequencies equal to the value of the transfer function of G at zero frequency. The model is used in conjunction with a function generator and a small sinusoidal signal to compute the penalty as the apparent operating point oscillates about the actual operating point. The 2-phase servomotor M is energized by a reference voltage in quadrature with the input to G', and its control phase is energized by the output of the function generator. The block marked "servo" is a conventional servomechanism which is used to convert the low-power variations in the angle of the servomotor shaft to the high-power variations in the input to the process.

In the 3-dimensional xyp space, the

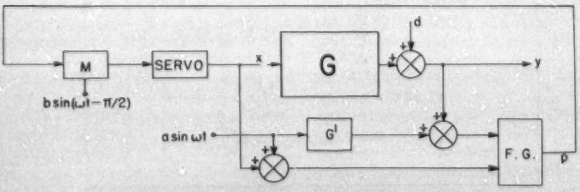


Fig. 11. Block diagram of complete single-variable regulator

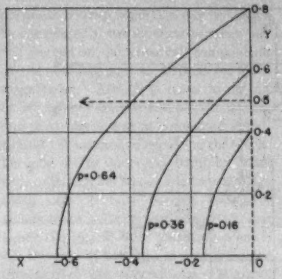


Fig. 13. Indifference curves corresponding to the same penalty function as in Fig. 12

plane $y=G.x+d$ intersects the surface $p(x,y)$ in a curve which gives the penalty as a function of x alone, for that particular value of d. Fig. 12 shows a typical set of such curves, with d as a parameter, when $G=1$ and $p(x,y)=|x|+y^2$. Fig. 13 shows the indifference map for the same system, for the same penalty function. The path of the minimum-penalty point as d increases is shown on each diagram by a dashed line. It can be seen in both diagrams that, as d increases from zero, the value of x remains at zero until $d=1/2$, and then increases linearly with further increase in d. If the penalty function is different the path of the operating point will, of course, be different.

Two-Input Two-Output Regulator

The block diagram of the 2-input 2-output regulator is shown in Fig. 14. As in the single-variable case, the values of the inputs and outputs are the incremental values. The equation for the regulator is

$$Y=HX+D \qquad (4)$$

where Y, X, and D are vectors with two components each, and H is a matrix transformation. The penalty is now a function of the four variables x_1, x_2, y_1, and y_2. For a given pair of disturbances d_1 and d_2, the penalty is a unique function of x_1 and x_2, i.e., the intersection of the hyperplane $Y=HX+D$ and the hypersurface $p(x_1,x_2,y_1,y_2)$ is a surface in the x_1x_2p space. It is convenient to think of this surface as an elastic membrane whose shape is distorted by variations in the values of the disturbances, thus shifting the co-ordinates, in the x_1x_2 plane, of the point of minimum penalty. Since there are two degrees of freedom in the system, it is necessary to adjust x_1 and x_2 to achieve the minimum value of penalty. Small-signal excursions of the apparent operating point can in this case be achieved by applying a 2-phase input (two signals in quadrature) to a low-power steady-state model of the basic system (H'), as shown in Fig. 15. This causes the apparent operating point to rotate around the actual operating point in the x_1x_2 plane, and once again the phase of the output of the function generator shows which direction the inputs should be altered. Two 2-phase servomotors M_1 and M_2 are used, the control phases of both being energized by the output of the function generator and each reference phase being energized by a voltage of the appropriate phase. This system will therefore adjust the inputs until the condition of minimum penalty is reached.

The remarks in the previous section, on the subject of the cost of altering the input in a single-variable regulator, may be applied with even greater significance to the 2-variable system. If the output side only is considered, the state of the system may be represented by a point on the y_1y_2 plane. It is possible to manipulate the two inputs in such a way as to move the output state from any point on this plane to any other point, provided that the transforms of the two unit vectors of the input are linearly independent. However, if these two vectors in the y_1y_2 plane are almost parallel, it may be necessary to use excessively large inputs in order to achieve a relatively small variation in the output state.

Tandem Cold-Rolling Mill

An excellent example of a multiactuator control system is the tandem rolling mill used in the manufacture of steel strip. The control of such a mill has already been described in considerable detail.[2,3] There are two independent variables associated with each stand in such a mill, namely, the nominal roll separation and the torque of the driving motor, so that the best scheme of control is by no means obvious. However, it may be assumed that satisfactory control of a 5-stand mill can be achieved by adjusting the roll separation of stand 1 (coarse control of output thickness) and the torque of the driving motor of stand 5 (fine control of output thickness).[2] Since both these adjustments have a significant effect on the tension between stands 4 and 5,[2] the system must be controlled in such a way that this tension stays within certain limits. In other words, the four variables on which the system penalty depends are: roll setting of stand 1, torque of stand 5, tension between stands 4 and 5, and thickness of output strip. It should therefore be possible to control the system in an optimum way with respect to these four variables, using the methods described here, if a suitable penalty function is used.

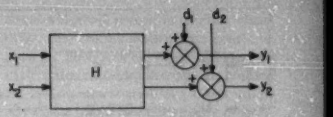


Fig. 14. The 2-input 2-output regulator

Penalty Function in Rolling Mill

Obviously, the penalty function must be related in some way to the increase in the cost of operation of the process when the system departs from its nominal state of operation. Therefore each of the significant variables will be examined in turn to find out what the nature of the penalty should be; the penalty functions given here are merely the suggestions of the author and are not intended to represent the best possible functions for such a situation.

Roll Setting of Stand 1

Because of the very large forces involved, alteration of the roll setting of stand 1 causes slight deformation of the rolls and therefore the transverse profile of the strip is slightly distorted. The deformation is probably directly pro-

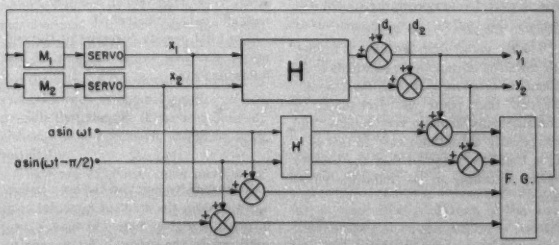


Fig. 15. Block diagram of 2-input 2-output regulator

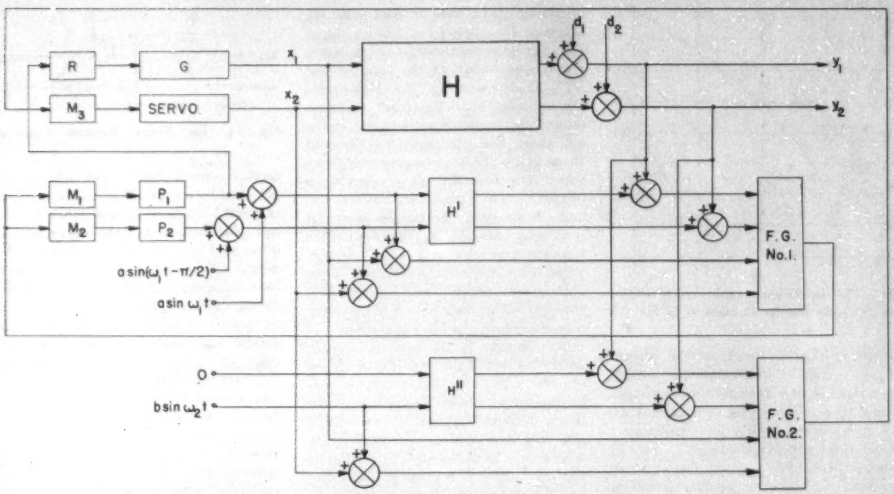


Fig. 16. Block diagram of 2-input 2-output regulator for tandem rolling mill, with on-off control on roll setting (x_1)

portional to the change in roll force, so that the penalty might be proportional to the absolute value of the change in roll setting, i.e.,

$$p(x_1) = a_1|x_1| \tag{5}$$

where x_1 is the incremental roll setting of stand 1 and a_1 is a constant.

Torque of Stand 5

For most efficient use of the mill, high rolling speeds are used and the motors develop almost their total rated power. Small variations in torque incur little or no penalty but larger variations are undesirable. Assume that the penalty associated with this variable is

$$\left.\begin{matrix} p(x_2) \\ |x_2| \leqslant b_2 \end{matrix}\right| = 0, \quad \left.\begin{matrix} p(x_2) \\ |x_2| > b_2 \end{matrix}\right| = a_2(|x_2| - b_2) \tag{6}$$

where x_2 is the incremental torque of stand 5, and a_2 and b_2 are constants.

Tension Between Stands 4 and 5

The danger of strip breakage is negligible until the tension between stands 4 and 5 has increased by a considerable amount. Beyond that point the probability of breakage may increase very rapidly, and the penalty should also vary in this way. It is therefore assumed that the penalty varies according to the following equation:

$$\left.\begin{matrix} p(y_1) \\ |y_1| \leqslant b_3 \end{matrix}\right| = 0, \quad \left.\begin{matrix} p(y_1) \\ |y_1| > b_3 \end{matrix}\right| = a_3(|y_1| - b_3)^3 \tag{7}$$

where y_1 is the incremental tension between stands 4 and 5, and a_3 and b_3 are constants.

Thickness of Output Strip

In practice the one thing which matters is that the strip should be within the commercial tolerance. This would imply that a suitable penalty function would be a step function at the appropriate thickness, but such a penalty would give no indication of how closely the thickness was approaching the boundary. Therefore it is better to choose a penalty function which places small emphasis on small errors but much larger emphasis on larger errors. A square-law penalty would probably be suitable, so the following penalty is proposed:

$$p(y_2) = a_4 y_2^2 \tag{8}$$

where y_2 is the incremental output thickness and a_4 is a constant.

The total penalty function for the mill might therefore be a function of the following form:

$$p = \begin{matrix} a_1|x_1| + a_2(|x_2| - b_2)^2 \\ + a_3(|y_1| - b_3)^3 \end{matrix} \left.\begin{matrix} \\ |y_1| > b_3 \end{matrix}\right| \left.\begin{matrix} |x_2| > b_2 \\ + a_4 y_2^2 \end{matrix}\right| \tag{9}$$

If the penalty function has been judiciously chosen, the total incremental cost incurred in a given period of time is proportional to the integral of the penalty function over that time. However, there are certain factors which may have to be considered, in which the method of finding the total incremental cost is not so obvious. One example of this will now be considered. It is realized that the example studied here, namely, the problem of wear in the screwdown mechanism, may not be so important in practice as is implied in this paper. However, the example serves to illustrate a principle which is likely to be valid in many practical situations, and therefore it is considered worth while to discuss it here.

Effects of Wear in the Screwdown Mechanism

In most rolling mills the roll setting of any stand is altered by means of a screwdown mechanism. If this mechanism is operated very frequently the wear of the nuts and screws may be excessive, due to the large forces involved. The mill should therefore be operated in such a way as to take account of this fact. It seems reasonable to assign a penalty to the movements of the screwdown mechanism, in such a way that the penalty integrated over a given period is proportional to the total distance traveled by the screws. This would mean that the instantaneous penalty should be proportional to the absolute value of the velocity. Then, since screwdown position and velocity would both be independent variables from the point of view of the penalty function, the finding of the

optimum control condition can no longer be considered a steady-state problem but will depend on the frequency components of the disturbances as well as on their magnitudes. Also, it is usual to use an on-off control on the screwdown motors and drive the screws at a constant speed, so that the method of exploring the neighborhood of the operating point by small sinusoidal signals would not be valid because only certain discrete values of steady-state velocity are permitted, and hence the steady-state operating point is not capable of continuous change in the direction of the velocity axis. If a proportional control were used, with continuously variable velocity, it seems likely that other difficulties would be encountered. For example, the presence of stiction would have a detrimental effect on the smoothness of control, and it is likely that stiction would also become a significant factor in the wear of the screws thus necessitating a further penalty based on the number of times the screws start to move.

The amount of screw operation can be reduced by retaining the on-off control and incorporating a suitable dead zone so that the screwdown mechanism does not operate unless a definite minimum amount of adjustment to the roll setting is required. This is also difficult to achieve by the use of small sinusoidal signals exploring the neighbourhood of the operating point, because it is not possible to find the co-ordinates of the minimum point from observations in a small region around the operating point. It is possible to overcome this difficulty by using a more elaborate system as shown in Fig. 16. This system uses two low-power steady-state models of the system, H' and H'', and two identical penalty function generators. The models H' and H'' are identical with each other, and each has the same transfer characteristics at all frequencies as the basic system has at zero frequency. The model H'', with an oscillating input in the x_2 direction only, operates in conjunction with function generator 2 and the 2-phase servomotor M_2 to maintain x_2 (the incremental torque of stand 5) at the value corresponding to minimum penalty, where minimum penalty in this case means the lowest value of penalty which can be achieved by altering x_2 if x_1 is fixed at its existing value. At the same time the model H', with a 2-phase input, operates in conjunction with function generator 1 and the two 2-phase servomotors M_1 and M_2 in such a way that the shaft positions of M_1 and M_2 correspond respectively to the discrepancies between the actual values of x_1 and x_2 and the values corresponding to the absolute minimum penalty (i.e., the lowest value of penalty which can be achieved by varying both x_1 and x_2). The potentiometers P_1 and P_2 are used to convert the shaft positions of the servomotors to corresponding voltages. The relay R and the screwdown motor G act like the forward loop of a servomechanism whose error quantity is the output voltage of P_1. The relay will operate when the discrepancy between the actual value and the desired value of x_1 is above a certain value. It can easily be arranged that the relay remains closed until the error voltage is zero, so that the screwdown mechanism will be driven until the optimum value is reached. In the meantime H'' and the associated control circuit maintain x_2 at the optimum value at all times, regardless of the value of x_1. It is obvious that this method of control will be effective in reducing the amount of screwdown movement, but of course the integral of the instantaneous penalty over a period will in general be greater than if the roll setting changed to correct each little variation. The best value of relay dead zone could be found by simulating the system and submitting it to external disturbances of magnitude and duration similar to those expected in practice, assuming that the total screw motion in a given period contributes to the integrated penalty over that period.

Practical Feasibility

The physical construction of control systems like those described in this paper would not be difficult. Referring to Figs. 15 and 16, it may be seen that most of the controller, including the 2-phase supplies, could be obtained by suitable interconnection of analog computer components. Almost any commercial analogue computer could be used for this purpose. Also, it would be possible to use digital equipment and sampled-data techniques to compute values of the penalty function at selected points in the neighborhood of the operating point, and to make logical decisions as to the best course to take. The cost of operation and the value of the product in industrial processes such as the rolling of steel are so high that the use of such equipment, although it is expensive, would be fully justified.

There are, of course, other methods of altering the system inputs than by 2-phase servomotors. For example, if the incremental output of the function generator is multiplied by the sinusoidal test signal, the product is a unidirectional signal which will reverse in polarity whenever the minimum-penalty point is passed. Such a signal could therefore be used (after suitable filtering) to control the system input variables. Other methods will be found in the literature describing optimizing or optimalizing control systems.

Although the methods outlined in this paper are valid only in the linear region of operation, it is possible to control a nonlinear system in a similar manner if the controller is made more elaborate. Such a controller would include a different kind of model of the physical process, a model which would copy the dynamic behavior and the nonlinearities of the actual process, rather than a steady-state model as in the controllers which are described in the paper.

Conclusions

1. This paper has attempted to show that one of the basic problems of system engineering is the same as one of the basic problems of economics, namely, the economic use of limited resources. Some of the relevant portions of economic theory have been outlined, and the application to control systems has been illustrated by considering the 2-actuator problem. The solution of the problem involves the use of a penalty function, the criterion for the choice of operating point being the minimization of penalty. To allow the problem to be solved automatically, information about the penalty is provided in the controller by a function generator.

2. Proceeding from the 2-actuator problem, the paper has given a method of controlling a 2-dimensional multivariable process such as a tandem rolling mill. The nature of a possible penalty function for such a mill has been considered, the penalty being based on both the input variable and the output variable.

3. The higher the standard of performance required from a process, the greater are the demands made on the total system in terms of power input, wear and tear on equipment, and the like. It is suggested that more thought should be given to the system as a whole, rather than concentrating on reaching the ultimate minimum error according to some criterion without reckoning the incremental cost of removing the last unit of error. This cost may be the cost of the input power or fuel or some other substance, or the cost of wear and tear on equipment, or it may be something quite different such as the increase in complexity as the perform

ance is improved. In the case of increasing system complexity, as in the other cases, there is usually a point beyond which it is uneconomical to go.

4. The problem of sharing the effort among several actuators or several processes is one which occurs in many fields, one of the most obvious to the electrical engineer being the division of load among several generators or among several power stations.[4] That particular problem is also amenable to automatic methods of solution.[5]

References

1. Value and Capital (book), J. R. Hicks. Oxford University Press, London, England, 1946, chaps. I, II.

2. Analysis of Tandem Cold Reduction Mill with Automatic Gauge Control, R. A. Phillips. *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 355–63.

3. Incremental Control Equations for Tandem Rolling Mills, J. H. Courcoulas, J. M. Ham. *Ibid.*, pp. 363–74.

4. Economy Loading of Power Plants and Electric Systems (book), M. J. Steinberg, T. H. Smith. John Wiley & Sons, Inc., New York, N. Y., 1943.

5. An Optimalizing Computer Controller for the Electric Utility Industry, L. K. Kirchmayer. *ASME Paper no. 58-IRD-2*, American Society of Mechanical Engineers, New York, N. Y., 1958.

Application of Continuous System Design Concepts to the Design of Sampled Data Systems

STANLEY F. SCHMIDT
ASSOCIATE MEMBER AIEE

THE THEORY associated with the design of sampled data systems[1] has progressed to the point where it is possible to design a digital controller pulse transfer function so as to obtain an overall system pulse transfer function which will meet certain specifications such as being stable, and physically realizable; having a minimum settling time in response to a given input; having ripple-free operation after a short transient, etc.

A survey of the literature, however, indicates a scarcity of information on the sensitivity of a system to parameter variations and on how to control, in the design process, the abruptness of the response. Barker[2] and Bertram[3] touch upon this subject and show that one method of smoothing the response is to design the characteristic equation so that some of the poles of the closed loop transfer function are located at points in the z-plane other than the origin. It is the purpose of this paper to show: first, that considerations of the sensitivity of the response to parameter variations and abruptness of the response to transient inputs preclude the use of finite settling time design (i.e., one in which the error is reduced to zero in a specified number of sampling instants); and second, to show that continuous system design concepts can be used to select the dominant poles of the closed loop pulse transfer function so as to achieve a desired transient response.

To obtain some indication of the practical limitations imposed on the response of the system, the "plant" is taken to be an aircraft and the system to be one which controls the aircraft's altitude. The abruptness and the sensitivity of the response to a transient maneuver will be judged in this application by the magnitude of the maximum load factor and the change in the stability of the system with parameter variations respectively.

System Description and Design Specifications

A block diagram of the altitude control system is shown in Fig. 1, where $r(t)$ and $c(t)$ are respectively the commanded and true altitude of the aircraft. The samplers are indicated by the symbol T. The transfer function used for the hold circuit and the aircraft autopilot, in Laplace transform notation is

$$G(s)=\left[\frac{1-e^{-sT}}{s}\right]\left[\frac{83.7(s+9.4)(s-8.8)}{s^2(s^2+4.2s+8.9)}\right] \quad (1)$$

The term within the first bracket in equation 1 is the mathematical expression for a zero-order hold circuit. The second term represents the aircraft-autopilot combination having a natural frequency of approximately 0.5 cycle per second and a damping ratio of 0.7. The s^2 term in the denominator arises from the double integration between normal acceleration and vertical distance or altitude.

The design problem, of course, is to select the least complicated digital controller, $D(z)$, to provide stability and other desired characteristics to the complete system. For this application these "other desired characteristics" are chosen to be:

1. A 50-foot step in altitude should not require a normal acceleration response exceeding the assumed 3 g maximum maneuvering capability of the aircraft.

2. The acceleration and altitude response should be insensitive to variations in the aerodynamic characteristics due to large changes in air density (differences in preset altitude); they should be relatively smooth (the sampling rate should be high); and they should be little affected by reasonable limits on control surface rates.

Unlike more conventional requirements of sampled data systems such as zero ripple or finite settling time, these specifications cannot be formulated into an equation for $D(z)$. One can, however, use conventional design methods and investigate the performance of the resulting system with regard to all the desired characteristics, and decide the extent to which the methods lead to a $D(z)$ design which will meet the specifications.

Finite Settling Time Designs

In general, the design equation for the digital controller pulse transfer function is

$$D(z)=\left(\frac{K(z)}{1-K(z)}\right)\left(\frac{1}{G(z)}\right) \quad (2)$$

To insure the the designed system to be stable, have zero ripple, zero steady-state error to some input transient, etc., specifications in mathematical terms are placed on $K(z)$, the closed loop pulse transfer function $C(z)/R(z)$, and $1-K(z)$, the transfer function relating $E(z)/R(z)$.

Paper 58-1083, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted September 2, 1958; made available for printing December 3, 1958.

Stanley F. Schmidt is with the National Aeronautics and Space Administration, Moffett Field Calif.

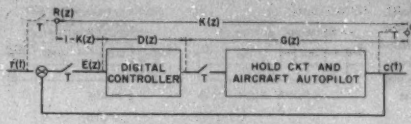


Fig. 1. Block diagram of an aircraft altitude control system

Now, it seems reasonable to assume that a finite settling time system would not be designed without imposing the zero ripple constraint, and of course, the system must be stable. Equations 3 and 4 indicate the mathematical form of the specification to meet only the zero ripple and stability criteria.

$K(z)$ = (numerator $G(z)$)(polynomial with undetermined coefficients)

$$= \left(\sum_0^J \alpha_k z^{-k}\right)\left(\sum_0^L a_l z^{-l}\right) \qquad (3)$$

$1-K(z)$ = (poles of $G(z)$ on or outside of the unit circle)(undetermined coefficients)

$$= \left(\sum_0^M \beta_k z^{-k}\right)\left(\sum_0^N b_n z^{-n}\right) \qquad (4)$$

Subtracting equation 3 from unity and equating coefficients of equal powers of $1/z$ with those of equation 4 allows the unique determination of the a_l's and the b_n's. There is a minimum L and N which will satisfy equation 3 and 4; if more are used then the additional a_l's and b_n's are not uniquely determined but may be chosen arbitrarily. If, however, more constraints are added, such as specifying the time response for several sampling instants, then a larger L and N will be required. Substitution of equations 3 and 4 into equation 2 gives the digital controller pulse transfer function as

$$D(z) = \frac{\left(\sum_0^J \alpha_k z^{-k}\right)[\text{stable poles of } G(z)]}{\sum_0^N b_n z^{-n}} \qquad (5)$$

One item to note from this equation is that any stable poles of $G(z)$ will occur as zeros of $D(z)$. Thus, this design process will always involve cancellation if $G(z)$ has some poles in the left half-plane. The design of $K(z)$ on the basis indicated in equations 3 and 4 results in a system with finite settling time, since $K(z)$ will always be a finite polynomial with a last term of $z^{-(J+L)}$ ($J+L$ is equal to the number of sampling instants necessary to correct a transient). Now, it is obvious that the sampling period T has not entered the design process except in the determination of $G(z)$. Thus, whether the system is operated at a sampling frequency of 10 cycles per second or of one cycle per second, the same number of sampling instants will be required to correct an input transient.

To illustrate the effects of sampling period on the altitude autopilot, a finite settling time design was made at two sampling periods, $T=0.25$ second and $T=1.0$ second. The pulse transfer functions for $G(z)$, $K(z)$, $1-K(z)$, and $D(z)$ for these two periods are given in the following equations 6 through 13.

For $T=0.25$ second:

$$G(z) = \frac{1.02z^{-1}-10.21z^{-2}-7.53z^{-3}+0.730z^{-4}}{(1-z^{-1})^2(1-1.02z^{-1}+0.349z^{-2})} \qquad (6)$$

$$K(z) = (1.02z^{-1}-10.21z^{-2}-7.53z^{-3}+0.730z^{-4})(-0.215+0.153z^{-1}) \qquad (7)$$

$$1-K(z) = (1-z^{-1})^2(1+2.22z^{-1}+1.085z^{-2}-0.112z^{-3}) \qquad (8)$$

$$D(z) = \frac{-0.215(1-0.710z^{-1})(1-1.02z^{-1}+0.349z^{-2})}{1+2.22z^{-1}+1.085z^{-2}-0.112z^{-3}} \qquad (9)$$

For $T=1.0$ second:

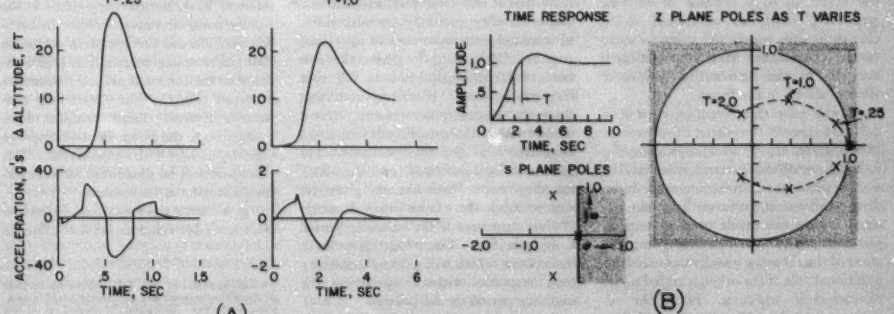
$$G(z) = \frac{-110.3z^{-1}-572.1z^{-2}-200.6z^{-3}-7.469z^{-4}}{(1-z^{-1})^2(1+0.125z^{-1}+0.0149z^{-2})} \qquad (10)$$

$$K(z) = (-110.3z^{-1}-572.1z^{-2}-200.6z^{-3}-7.469z^{-4})(-0.0035+0.00238z^{-1}) \qquad (11)$$

$$1-K(z) = (1-z^{-1})^2(1+1.614z^{-1}+0.487z^{-2}+0.0178z^{-3}) \qquad (12)$$

$$D(z) = \frac{(-0.0035)(1-0.679z^{-1})\times(1+0.125z^{-1}+0.0149z^{-2})}{(1+1.614z^{-1}+0.487z^{-2}+0.0178z^{-3})} \qquad (13)$$

The digital controller for both systems was simulated on an analog computer with the use of the method described in reference 4. The transient time responses of both altitude and normal acceleration for the two systems are shown in Fig. 2(A). Of particular note is the large overshoot of the altitude response for both sampling periods, which is a characteristic of a finite settling time design. Of greater importance, however, is the



Figs. 2. Effects of sampling period T

A—On transient response

B—On pole positions

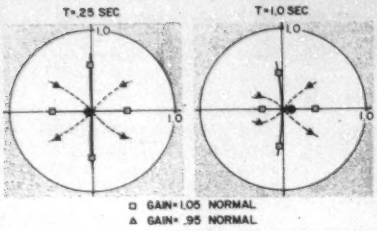


Fig. 3 (left). Root locus plot in z-plane illustrating sensitivity of finite settling time designs

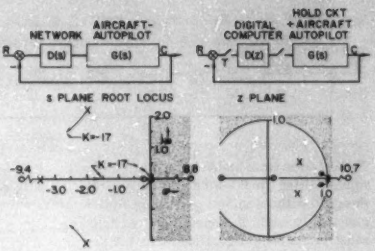


Fig. 4 (right). Design of altitude control system in s plane and corresponding z-plane problem

fact that the peak magnitudes of the normal acceleration differ by a factor of approximately 30. The peak of 32 g's for the $T=0.25$ case obviously could not be achieved by a conventional aircraft and is only a consequence of the linear analysis. For the $T=1.0$ second example the maximum acceleration is slightly higher than is listed in the specification (3 g's for a 50-foot step). These data, however, illustrate that if the sampling period were a variable and could be made somewhat larger than one second, this design could satisfy the abruptness specification.

Another question of importance is the sensitivity of the finite settling time design system to parameter variations. Unpublished simulation studies have indicated that such systems are very sensitive to variations of the aerodynamic parameters. Some appreciation of the sensitivity can be gained with reference to the root locus plot for the two systems shown in Fig. 3 where it can be noted that both systems are very sensitive to small ($\pm 5\%$) open-loop gain changes around the design value of gain (which places all the poles, except those cancelled, at the origin). The root locus plot illustrates the origin to be a position of extreme sensitivity since an additional $\pm 5\%$ change in gain causes the poles to move the short distance from the indicated symbols to the terminating points of the loci shown in the figure.

As was previously mentioned, if it is desired to specify the output time response (for a given input) at chosen sampling instants, additional constraints on the system must be taken into account in the design process of finite settling time designs. This will result in a more complicated $D(z)$. For each sampling instant of the response which is specified, an additional pole at the origin is added to the characteristic equation. From the results presented in Fig. 3 one could conclude that adding more poles will increase sensitivity. Thus, although the abruptness of the response can be controlled by adding more constraints if the sampling period is short, or by increasing the sampling period if it is a variable, the sensitivity of the finite settling time designs to parameter variations will remain large.

System Design Using Continuous System Design Concepts

As has been shown, finite settling time designs are extremely sensitive to parameter changes, due principally to placing the uncancelled poles at the origin in the z-plane. This would be true also if the characteristic equation is of the form $(1-cz^{-1})^n$ as had been proposed by Barker.[3] What appears to be needed, then, is some way of selecting pole locations in the z-plane not all at the same point and which correspond to some known time response that will meet the specifications.

Fig. 2(B) indicates one of the difficulties associated with designing in the z-plane and how it may be overcome by working in the s-plane. Consider the time response shown and presuppose that this response will neet the set specifications in terms of speed and abruptness. Note that it is not too fast, and has adequate stability (actually, corresponding to a second-order system with $\omega_n=1$ and $\zeta=0.5$). Now, in the s-plane, the pole (and zero) correspondence to the time response is fixed. However, note that this is not the case in the z-plane. Here it is shown that the sampling period has a large effect on the pole positions. For each pole location, then (and the corresponding zeros which are not given for convenience), there is an infinite number of time responses if the sampling period is not specified. Thus, locating poles in the z-plane, which will give a satisfactory time response without specifying the sampling period, is not possible.

One possibility of overcoming this difficulty is suggested from a study of Fig. 2(B); i.e., to design, first, a system in the s-plane, transfer the dominant s-plane poles at the system operating sampling period to the z-plane, and then proceed with the design of the digital controller in the z-domain. The merit of this approach is given some support by noting that a sampled data system gives exactly the same response as the continuous system from which it is derived if the sampling period T approaches zero. This may be seen by considering the series expansion of z.

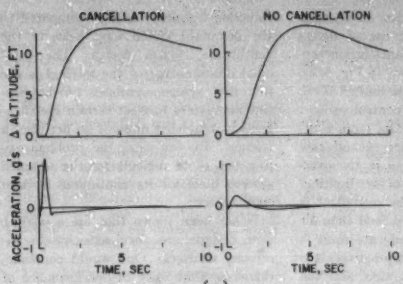
$$z = e^{sT} = 1 + sT + (sT)^2 + \ldots \qquad (14)$$

A mapping in the s-plane in the vicinity of the origin corresponds to a mapping in the z-plane around $z=1$, except for the scale factor T. As T approaches infinity, all stable poles in the s-plane move toward the origin of the z-plane and thus a system designed by fixing the poles in the s-plane approaches a finite settling time design. In between these two extremes there is a large range of sampling periods where the response of the sampled data system is, for all practical purposes, identical with the continuous system.

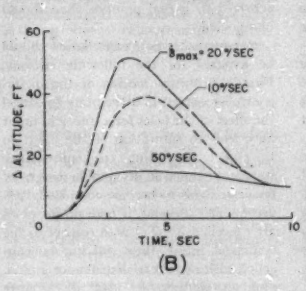
The proposed procedure, then, for designing a sampled data system is as follows: First, design the system in the s-plane using continuous system methods. Second,* choose the sampling period so that the sampling frequency is high compared to the dominant natural frequencies or break points of the closed loop continuous system. Third, transfer dominant poles to the z-domain and design the digital controller with z-transforms. This procedure will be illustrated with the aircraft-autopilot as the example.

Fig. 4 shows the root locus of the aircraft-autopilot example for a continuous

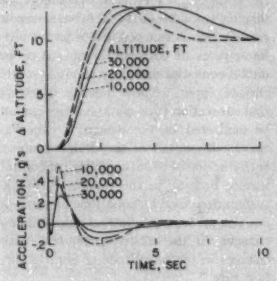
* There appears, at this time, to be no method of specifying numerically the lowest possible sampling frequency. Sampling, of course, gives rise to error. Unpublished data for the aircraft altitude control system designed for sampling frequencies, approximately 12 and 48 times the dominant continuous system second-order mode frequency, show approximately the same response to the same input.



(A)



(B)



(C)

Fig. 5. Step response showing various effects

A—Effects of pole cancellation
B—Effect of control surface rate limits; pole cancellation permitted
C—Effect of change in basic or preset altitude; no pole cancellation permitted

system with a particular choice of lead network

$$D(z) = -\left(\frac{z+0.25}{z+3.5}\right) \times \text{constant}$$

It should be noted that the simplest possible stabilizing network for this example is a lead network and that the aircraft-autopilot zero in the right half-plane forces the system gain to be negative for stability. The design in the s-plane is deliberately made sluggish (the dominant poles are kept close to the origin compared to the variable poles due to aircraft dynamics) to reduce the sensitivity of the system to changes in the aircraft parameters. For the design gain of -17, the dominant poles (poles nearest the origin) are at $\omega_n = 0.5$ and $\zeta = 0.7$. The open-loop s-plane poles and zeros are given also in Fig. 4 for a sampling period of 0.25 second. The dots in the z-plane show the desired location of the dominant closed loop poles mapped from the s-plane and the dotted lines indicate the approximate loci these poles will take when the gain of the closed loop system is varied from zero to the desired operating point. Note that, as was true for the continuous system, the zero at $z = 10.7$ requires that the gain for the sampled data system must be negative for stability.

Once the desired dominant mode poles and zeros are specified in the z-plane (for the system-sampling rate) one can use a trial-and-error procedure with the root locus method to find the digital controller pulse transfer function and gain which will place the open-loop poles at these positions. However, an analytical method for accomplishing the same purpose may be derived in a manner similar to that used by others in obtaining equations 3, 4, and 5 for the finite settling time design. The details of the derivation will not be given here, but the necessary equations and the philosophy for two possible approaches to the design of the digital controller pulse transfer function, $D(z)$ follow. For example, if desired, one can design $D(z)$ to have two complex zeros to cancel the two complex poles of $G(z)$ and place other poles and zeros of $D(z)$ at points in the z-plane in such a way that all poles of the closed loop pulse transfer function $K(z)$ are at the origin, except those cancelled and those of the desired dominant mode. A second method is to design $D(z)$ so that no pole cancellation is permitted. If pole cancellation technique is used, equations 15 and 16 indicate the manner in which the desired constraints are placed on $K(z)$ and $1\text{-}K(z)$

$$K(z) = \frac{[\text{Numerator of } G(z)](\text{undetermined coefficients})}{\text{desired dominant mode}} = \frac{\left(\sum_0^J \alpha_k z^{-k}\right)\left(\sum_0^L a_l z^{-l}\right)}{1 - 1.825z^{-1} + 0.84z^{-2}} \quad (15)$$

$$1 - K(z) = \frac{(G(z) \text{ poles on or outside the unit circle})(\text{undetermined coefficients})}{\text{desired dominant mode}} = \frac{\left(\sum_0^M \beta_m z^{-m}\right)\left(\sum_0^N b_n z^{-n}\right)}{1 - 1.825z^{-1} + 0.84z^{-2}} \quad (16)$$

The desired dominant mode given in the foregoing equations (a second-order transfer function with $\omega_n = 0.5$ and $\zeta = 0.7$ that has been transferred to the z-domain for $T = 0.25$) is indicated by the two dots on the z-plane plot of Fig. 4. If pole cancellation is not permitted, equations 17 and 18 replace equations 15 and 16 respectively.

$$K(z) = \frac{[\text{Numerator of } G(z)] \times (\text{undetermined coefficients})}{(\text{desired dominant mode}) \times (\text{undetermined coefficients})} = \frac{\left(\sum_0^J \alpha_k z^{-k}\right)(a_0 + a_1 z^{-1})}{1 - 1.825z^{-1} + 0.84z^{-2}) \times (1 + c_1 z^{-1} + c_2 z^{-2})} \quad (17)$$

$$1 - K(z) = \frac{[\text{Denominator of } G(z)] \times (\text{undetermined coefficients})}{(\text{desired dominant mode}) \times (\text{undetermined coefficients})} = \frac{\left(\sum_0^P \beta_p z^{-p}\right)(b_0 + b_1 z^{-1})}{(1 - 1.825z^{-1} + 0.84z^{-2})(1 + c_1 z^{-1} + c_2 z^{-2})} \quad (18)$$

For this example the substitution of equations 17 and 18 into equation 2 shows that

$$D(z) = \frac{a_0 + a_1 z^{-1}}{b_0 + b_1 z^{-1}} \quad (19)$$

An even simpler $D(z)$ can be obtained if a third undetermined coefficient is added to the denominator of equations 17 and 18. In this case it will be found that b_1 can be chosen to be zero and $K(z)$ will have no poles at the origin. $D(z)$ then, will be of the form

$$D(z) = a_0 + a_1 z^{-1} \quad (20)$$

It is interesting to note that in both $D(z)$ designs in which pole cancellation is not permitted (equations 19 and 20) the digital computer pulse transfer function

has been simplified by adding a denominator to $K(z)$. Along this line the first work by Bertram[3] showed only that $D(z)$ need not be more complex if a denominator is added in terms of the delay line synthesis of the digital controller he was considering.

Equations 21 through 26 give the numerical values of the aircraft-autopilot example at a sampling period of 0.25 second for the two design methods indicated by equations 15 through 18.

With cancellation permitted

$$K(z)=\frac{(1.02z^{-1}-10.21z^{-2}-7.53z^{-3}+0.730z^{-4})(-0.0122+0.0113z^{-1})}{1-1.825z^{-1}+0.84z^{-2}} \tag{21}$$

$$1-K(z)=\frac{(1-z^{-1})^2(1+0.187z^{-1}+0.0777z^{-2}-0.0083z^{-3})}{1-1.825z^{-1}+0.84z^{-2}} \tag{22}$$

$$D(z)=\frac{-0.0122(1-0.926z^{-1})\times(1-1.02z^{-1}+0.349z^{-2})}{1+0.187z^{-1}+0.0777z^{-2}-0.0083z^{-3}} \tag{23}$$

With cancellation not permitted

$$K(z)=\frac{(1.02z^{-1}-10.21z^{-2}-7.53z^{-3}+0.730z^{-4})(-0.00276+0.00259z^{-1})}{(1-1.825z^{-1}+0.84z^{-2})\times(1-1.205z^{-1}+0.402z^{-2})} \tag{24}$$

$$1-K(z)=\frac{(1-z^{-1})^2(1-1.02z^{-1}+0.349z^{-2})(1-0.0054z^{-1})}{(1-1.825z^{-1}+0.84z^{-2})\times(1-1.205z^{-1}+0.402z^{-2})} \tag{25}$$

$$D(z)=-0.00276\,\frac{(1-0.936z^{-1})}{1-0.0054z^{-1}} \tag{26}$$

The simplest possible $D(z)$ (equation 20) has been computed but no simulation data are available for this design. However, a comparison of the magnitudes of the coefficients of $D(z)$ indicate that there would be little difference in the response between that for the simplest possible $D(z)$ and the one given by equation 26.

Fig. 5(A) compares responses for the two systems (equations 23 and 26). Note that there is little difference between the two altitude time responses. However, the acceleration required if pole cancellation is permitted is considerably higher during the initial transient than when pole cancellation is not permitted. This difference in the acceleration response is associated with the fact that the digital computer's output is quite abrupt initially in cancelling the effects of the two poles. From these results it is concluded that little is gained by pole cancellation in terms of response time and one of its poor features is that a more complicated digital controller is required.

Fig. 5(B) illustrates an even more serious drawback to the use of design methods which employ cancellation procedures. These data show, in Fig. 5(B), the effect of having a rate-limited servo driving the aerodynamic control surface. Of particular interest is the large overshoot for a 20 degrees per second rate limit and the large change in the overshoot with the magnitude of rate limiting. The unusual fact that the overshoot is higher for 20 degrees per second than for the 10 degrees per second rate limit is due to the time relationship between the actual control surface motion and the command input during the initial part of the transient. The large overshoots are directly traceable to the fact that true pole cancellation is not possible where a nonlinearity exists between the output of the digital computer and the aircraft's control surface input. These results illustrate that saturation-type nonlinearities cannot be neglected in the design if cancellation is attempted in the design process. Introduction of the same values of control surface limiting in the simulation of the system described by equations 24 through 26 (no pole cancellation) showed no changes in the altitude response with changes in the magnitude of control surface rate.

Fig. 5(C) shows the effect of altitude change on the transient response. Note that in obtaining these data only the aircraft parameters were varied in accordance with the manner in which they were affected by a change in air density. A change from the design altitude, 30,000 feet to 10,000 feet results in a 40% increase in natural frequency of the aircraft and a 90% increase in gain. As may be seen, the response becomes faster as the altitude decreases due to the increase in aerodynamic gain but that no large change occurs in the stability of the system.

Conclusions

The design of a sampled data system in which considerations of the sensitivity of the response to parameter variations and the abruptness of the response are important has been shown to be simplified considerably by first designing a continuous system in the s-plane which will satisfy the requirements. The dominant characteristics of the continuous system are then transformed to the z-domain for final synthesis of the sampled data system. The principal advantage of this method is that the sampling period T has little influence on the transient performance over a very large range of values, provided the sampling frequency is high compared to the dominant closed loop poles of the continuous system design. The principal disadvantage of the method is the fact that system synthesis of the continuous system to meet certain specifications is somewhat more of an art than a science. In any case, the problem appears no more difficult for the sampled system then for its continuous system counterpart.

It has been shown that finite settling time designs are very sensitive to parameter changes. One would conclude, therefore, that since the performance of such sampled data systems cannot be achieved in actual practice they should be avoided.

Designs using pole cancellation should be avoided for the following reasons: First, the abrupt motion of the digital controller output in attempting to cancel the effect of the poles for a transient input may lead to saturation of the input to the plant. Saturating type nonlinearities frequently cause an acceptable design for linear analysis to become completely unacceptable. Second, if undesirable poles are cancelled by $D(z)$ with respect to the command input, they will not be cancelled with respect to distrubance inputs, and an undesirable load disturbance response may result.

References

1. Sampled Data Control Systems, **J. R. Ragazzini, G. F. Franklin.** McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

2. The Theory of Pulse-Monitored Servos and Their Use for Prediction, **R. H. Barker.** *Report no. 1046*, Signals Research and Development Establishment, Christchurch, England, Nov. 1950.

3. Factors in the Design of Digital Controllers for Sampled-Data Feedback Systems, **John E. Betram.** *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 151–59.

4. Network Compensation of Error-Sampled Feedback Control Systems, **J. Sklansky.** *Technical Report T-7/B*, Columbia University Press, New York, N. Y., April 1, 1955.

Discussion

Gene Franklin (Stanford University, Stanford, Calif.): Many of the early papers on sampled data system were concerned with extending the continuous system design methods of steady-state frequency-response shaping to apply to the new area. Later efforts tend to emphasize the differences between continuous and sampled data design and have developed new time domain design objectives such as finite settling time designs. This paper represents, in a certain sense, a return to the original objective of relating sampled data system design to continuous system design. In this paper,

however, the bridge is built from the time domain synthesis methods of the z-transform to the pole-zero design methods of continuous systems. There seems to be reason to expect that both design problems will benefit from the connection.

The primary purpose of this discussion is to drive an analytical nail in the coffin of the finite settling time design alongside the author's experimental fastening. Actually, despite the calculations to follow, the finite settling time design will remain a useful concept to illustrate time domain design methods and will probably remain a practical design for low-order systems which must operate with very long sampling periods.

The sensitivity of a system may, in general, be defined as the percentage change in the performance of the system to a percentage change in a parameter. That is, if P is the performance index, then the sensitivity of the system with respect to a parameter k is defined as

$$S_k^P = \frac{dP/P}{dk/k}$$

In the past, in connection with feedback amplifier and, later, in arbitrary feedback system, the performance P has been taken to the output/input gain ratio at some fixed frequency, usually zero (direct current). There is no reason why this should be so. Recently Huang[1] defined the "performance" as the location of the pole of the output/input transfer function in the s-plane or in the z-plane. Huang's definition and similar calculations were developed independently by the discusser using the root locus as a starting point. From this point of view, the characteristic equation is written in the form

$$1+kG(z)_t=0 \tag{26}$$

where k is the parameter of interest. The roots of this equation, $z=z_t$ are the poles of the output/input transfer function of the system. The shift in a particular pole position for a change in the parameter k may be obtained by setting the total derivative of equation 26 to zero.

$$dkG(z_t)+kG'(z_t)dz_t=0$$

$$\frac{dz_t}{dk}=-\frac{G(z_t)}{kG'(z_t)} \tag{27}$$

where the conventional notation

$$G'(z_t)=\left.\frac{dG}{dz}\right|_{z=z_t}$$

is used. Because of equation 1, it is also possible to write equation 27 in the forms

$$\frac{dz_t}{dk}=\frac{G^2(z_t)}{G'(z_t)}=\frac{1}{k^2G'(z_t)} \tag{28}$$

It should be realized that equations 27 and 28 represent the differential change in a root position for a differential change in the parameter k. To obtain percentage changes one needs to normalize the differential changes with respect to the sizes of the quantities involved. In this case, one is led to the definition

$$S_k^{z_t}=\frac{\dfrac{dz_t}{z_t}}{\dfrac{dk}{k}}=\frac{G(z_t)}{z_tG'(z_t)}=\frac{1}{z_tkG'(z_t)} \tag{29}$$

The sensitivity defined by equation 29 has many implications, only one of which will be considered here, namely that $S_k^{z_t}$ is infinite when $G'(z_t)$ is zero. Points where $G'(z)$ is zero are called saddle points of the complex plane representation of $G(z)$ and they include all points of multiple roots of the equation $1+kG(z_t)=0$. In other words, the finite settling time design which places all poles at the origin of the z-plane is operating at a point of infinite sensitivity of root position to parameter change! A design which shifts the multiple roots from the origin to a point such as $z=a$ does not change the sensitivity picture. An analysis of equation 29 shows that the speed with which the sensitivity decreases from infinity, as operation shifts from the saddle point, depends upon the order of the saddle point. The larger the number of poles placed at a single operating point, the more sensitive will be the root locations to parameter changes. It is largely for this reason that the low-order systems designed for finite settling time in the past have been relatively successful. The higher order, more practical system described in this paper shows the extreme sensitivity of the design.

REFERENCE

1. THE SENSITIVITY OF THE POLES OF LINEAR, CLOSED-LOOP SYSTEMS, R. Y. Huang. *AIEE Transactions*, vol. 77, pt. II, Sept. 1958, pp. 182–87.

Stanley F. Schmidt: The author appreciates Dr. Franklin's helpful suggestions given during the progress of this investigation. His derivation showing "finite settling time designs" to have infinite sensitivity at their operating conditions, gives further proof as to why they should be avoided and why a design approach such as is presented in the paper should be used.

The definition of sensitivity given by equation 29 has certain implications which should be of use to the control system designer; i.e., it gives a percentage shift in a pole position for a percentage change in some parameter. It is the author's opinion, however, that this does not completely resolve the question of the sensitivity of a closed loop system, since one usually desires to know the magnitude in the pole position shift for a specified incremental change in a parameter. At present, this information can only be obtained by the relatively tedious root locus method. Furthermore, the change in the pole position which is tolerable in a given control system depends on its position in the s-plane (or z-plane) and whether it is a "dominant" pole or one which is significantly far away from the origin in the s-plane (or unity in the z-plane) so that its position has a negligible influence on the desired response of the system.

From these standpoints, further research would be desirable to determine what additional definitions need be added to the sensitivity question as well as how to design control systems which have only a small sensitivity (or change in response) with parameter variations.

Electric and Magnetic Conditions Inside an Induction-Heated Workpiece

CHESTER A. TUDBURY
MEMBER AIEE

This paper deals with the case of a long, round, solid bar of homogeneous conducting material located coaxially inside a solenoidal coil carrying sinusoidal alternating current. This is a configuration used frequently in industrial induction heating installations. The same principles apply to more complicated arrangements. The results given here apply over the entire range of sizes and frequencies used in industrial induction heating, as well as to the lower frequencies and smaller dimensions employed to avoid

THE PURPOSE of this paper is to improve the general understanding of induction heating through a familiarity with what happens electrically and magnetically inside a solid bar being heated. AIEE Standard No. 54[1] defines induction heating as the heating of a nominally conducting material in a varying magnetic field due to its internal losses. This paper reports the results of a study of how these losses occur, and describes the vector current densities and flux densities inside the material being heated. The AIEE definition is sufficiently general to include not only intentional induction heating as practiced industrially for hardening, forge heating, brazing, and other applications, but also the eddy current and hysteresis heating which appear as core losses in motors and transformers.

Paper 59-64, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 6, 1958; made available for printing November 25, 1958.

CHESTER A. TUDBURY is with the New Rochelle Tool Corporation, New Rochelle, N. Y.

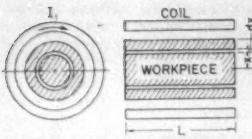


Fig. 1. Coil and workpiece

eddy currents and reduce core losses in electromagnetic equipment.

Accurate methods of designing induction heating coils and of calculating their electrical performance in advance have been available since the 1930's and are still being improved.[2,3] Their emphasis is more on the integrated effect of conditions inside the bar on the electrical characteristics at the coil terminals than on a study of these internal conditions per se.

Conditions

Fig. 1 shows the case under consideration. The bar and the coil are very long, so end effects are neglected. The magnetic field is parallel to the centerline of the bar. The primary current in the coil and the induced currents in the bar follow coaxial paths around the same center line.

Electric resistivity ρ and magnetic relative permeability μ are the two most important properties of the material of the bar. Resistivity depends upon temperature, which in turn depends upon internal power distribution (always greatest at the surface), specific heat, density thermal conductivity, rate of heating, and surface thermal losses. Permeability depends upon degree of magnetic saturation and also upon whether the temperature is below or above the Curie value. In most induction-heating applications for heating steel, the surface flux density is above saturation.

Hysteresis loss is also present when heating magnetic materials. It is insignificant relative to eddy current power in most industrial induction-heating applications.

A complete treatment of the foregoing is beyond the scope of this paper. Resistivity and permeability are assumed to be uniform and hysteresis losses are neglected.

Nomenclature

a = workpiece radius, cm (centimeters)
B = flux density, gausses
ber = Bessel function, real
bei = Bessel function, imaginary
ber′ = d, ber
bei′ = d, bei
c = reference depth, cm
E_2 = voltage induced in workpiece, rms, volts
f = frequency, cycles per second
H = magnetic intensity, oersteds
H_g = magnetic intensity in air gap, oersteds
H_r = real component of H
H_i = imaginary component of H
I_1 = current in coil, rms, amp (amperes)
I_2 = total current in workpiece, rms, amp
I_{2r} = real component of I_2
I_{2i} = imaginary component of I_2
K_R = resistance factor
K_x = reactance factor
L = length of coil and work, cm
N_1 = number of turns in primary coil
P_2 = total power developed in workpiece, watts
S = current density in workpiece, amp/sq cm (amp per square cm)
S_r = real component of S, amp/sq cm
S_i = imaginary component of S, amp/sq cm
Φ_2 = total flux in workpiece, lines
ρ = resistivity, ohm-cm
μ = relative permeability

Qualitative Explanation

Consider the bar to be composed of many thin coaxial sleeves. The magnetic intensity at the surface ($x=a$) is identical with the air gap intensity H_g, where

$$H_g = \frac{4\pi N_1 I_1}{10L}$$

The flux density in the outermost sleeve is H_g. The total flux surrounded by this sleeve induces a voltage in the sleeve. As a consequence, current flows circumferentially in the sleeve. The magnitude of this current is determined by the induced voltage and the resistance of the sleeve. This current, flowing in a substantially opposite instantaneous direction of the primary current, causes the magnetic intensity at the inner surface of the sleeve to be less than at the outer surface. The induced voltage is 90 electrical degrees out of phase with the flux which produces it, but the effect which the resulting current exerts upon that same flux is in phase with the current. The resulting flux density, induced voltage, and current in the second sleeve are thus not only smaller in magnitude but are also more lagging in time phase than in the outer sleeve.

The current in each sleeve is accompanied by an I^2R loss. The total power is the integration of the losses in all the sleeves. This power is densest at the surface of the bar and is zero at the center.

Mathematical Background

Differential equations for magnetic intensity H, and current density S, as functions of distance from the center line x, are obtained by letting the thickness of the sleeves approach zero, equating the induced voltage in a sleeve to its resistance drop, and taking into account the effect of the induced current on the magnetic intensity inside the sleeve.

$$\frac{d^2H}{dx^2} + \frac{1}{x}\frac{dH}{dx} - j\frac{1}{c^2}H = 0 \qquad (1)$$

where

$$C = \frac{10^9 \rho}{8\pi^2 \mu f}$$

$$S = \frac{10}{4\pi}\frac{dH}{dx} \qquad (2)$$

The boundary conditions are

At $x=a$, $H=H_g$.
At $x=0$, H is not infinity.
At $x=0$, $S=0$.

Solving equations 1 and 2, and applying these conditions yields

$$B = \mu H_g (H_r + jH_i) \qquad (3)$$

where

$$H_r = \frac{\text{ber } x/c \text{ ber } a/c + bei\ x/c \text{ bei } a/c}{\text{ber}^2 a/c + \text{bei}^2 a/c} \qquad (4)$$

and

$$H_i = \frac{\text{bei } x/c \text{ ber } a/c - \text{ber } x/c \text{ bei } a/c}{\text{ber}^2 a/c + \text{bei}^2 a/c} \qquad (5)$$

$$S = -\frac{N_1 I_1}{CL}(S_r + jS_i) \qquad (6)$$

where

$$S_r = \frac{\text{ber } a/c \text{ ber}' x/c + \text{bei } a/c \text{ bei}' x/c}{\text{ber}^2 a/c + \text{bei}^2 a/c} \qquad (7)$$

and

$$S_i = \frac{\text{ber } a/c \text{ bei}' x/c - \text{bei } a/c \text{ ber}' x/c}{\text{ber}^2 a/c + \text{bei}^2 a/c} \qquad (8)$$

The total flux in the bar is found by integrating equation 3 over the area of the bar:

$$\phi_2 = \mu H_g (2\pi ac)\left(\frac{K_x - jK_r}{\sqrt{2}}\right) \qquad (9)$$

where

$$K_x = \sqrt{2}\left(\frac{\text{ber } a/c \text{ bei}' a/c - \text{bei } a/c \text{ ber}' a/c}{\text{ber}^2 a/c + \text{bei}^2 a/c}\right) \qquad (10)$$

and

$$K_R = \sqrt{2}\left(\frac{\text{ber } a/c \text{ ber}' a/c + \text{bei } a/c \text{ bei}' a/c}{\text{ber}^2 a/c + \text{bei}^2 a/c}\right) \qquad (11)$$

This total flux generates a net back electromotive force E_2 in the bar:

$$E_2 = -\frac{\rho(2\pi a)}{cL} N_1 I_1 \left(\frac{K_R + jK_x}{\sqrt{2}}\right) \qquad (12)$$

The total current induced is found by integrating equation 6 from $x=0$ to $x=a$:

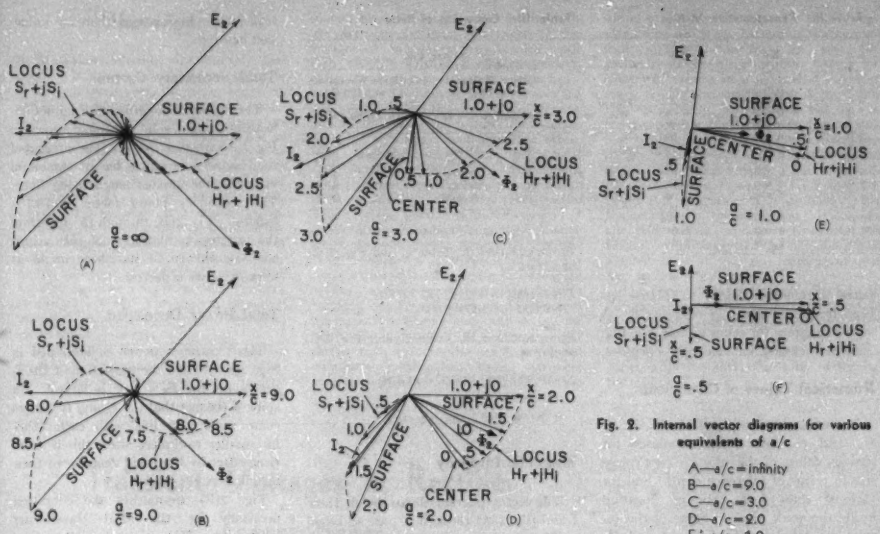


Fig. 2. Internal vector diagrams for various equivalents of a/c

A—a/c = infinity
B—a/c = 9.0
C—a/c = 3.0
D—a/c = 2.0
E—a/c = 1.0
F—a/c = 0.5

$$I_2 = N_1 I_1 (I_{2r} + jI_{2i}) \qquad (13)$$

where

$$I_{2r} = \left(\frac{\text{ber } a/c}{\text{ber}^2 a/c + \text{bei}^2 a/c} - 1 \right) \qquad (14)$$

and

$$I_{2i} = -\left(\frac{\text{bei } a/c}{\text{ber}^2 a/c + \text{bei}^2 a/c} \right) \qquad (15)$$

The total power developed by eddy currents inside the bar is found by integrating the I^2R in all the sleeves. It is also equal to

$$E_2 I_2 \cos \angle \frac{E_2}{I_2}$$

$$P_2 = N_1^2 I_1^2 \frac{\rho (2\pi a K_r)}{\sqrt{2} cL} \qquad (16)$$

Reference Depth

The quantity c, which first appeared in the basic differential equation 1 is of great significance; c has the dimension of length

$$c = \sqrt{\frac{10^9 \rho}{8\pi^2 \mu f}} = 3.560 \sqrt{\frac{\rho}{\mu f}} \text{ cm} \qquad (17)$$

In equation 16 the total power is the same as if a uniformly distributed current $N_1 I_1$ were flowing circumferentially around a cylinder of average radius $a K_r$, wall thickness of $c\sqrt{2}$, and length L. Various terms have been applied to the quantity c (also to $\sqrt{2}c$) by different authors: "skin depth," "reference dimension," "reference depth." "Reference depth" is used herein to mean the dimension c.

The key to the choice of frequency is the ratio of radius to reference depth. This ratio is the dimensionless parameter in all the Bessel functions in equations 3 through 15. It increases with frequency and with workpiece size. When $a/c = \infty$, K_R and K_X both equal 1.0. When $a/c = 0$, K_R and K_X are zero.

There is nothing in the definition of reference depth which implies smallness. Cases wherein the reference depth approaches or even exceeds the radius of the workpiece are of more than simply academic interest. The presently increasing use of line frequencies for induction heating, and the use of thin laminations in iron cores are examples. This is why the term "reference depth" is preferred over "skin depth." Equation 16 applies even when c exceeds a. However, it is admittedly difficult to visualize a hollow cylinder whose wall thickness exceeds its outer radius!

Infinite Radius: Flat Slab

If the radius of the round bar is increased to infinity the bar becomes a flat slab. The independent variable x, which is the distance from the center, loses its significance, and must be replaced by the distance below the surface. The first derivative in equation 1 disappears and the solution becomes simple exponential functions. The power is the same as if a uniform current $N_1 I_1$ flowed in a surface layer $\sqrt{2}\,c$ deep. This affords an easy way of obtaining approximate results for irregularly shaped workpieces where the cross-sectional dimensions are large com-

Table I. Representative Values

a/c	x/c	H_r	H_i	S_r	S_i
9.0	9.0	1.000	0.0	−0.920	−1.000
9.0	8.5	0.679	−0.251	−0.869	−0.450
9.0	8.0	0.398	−0.341	−0.702	−0.088
9.0	7.0	0.042	−0.273	−0.311	0.202
9.0	3.0	−0.078	−0.125	−0.056	0.188
9.0	5.0	−0.076	−0.024	0.040	0.097
9.0	4.0	−0.041	0.022	0.051	0.026
9.0	3.0	−0.005	0.024	0.032	−0.006
9.0	2.0	0.005	0.015	0.014	−0.013
9.0	1.0	0.011	0.007	0.004	−0.008
9.0	0.0	0.012	0.004	0.0	0.0
3.0	3.0	1.000	0.0	−0.764	−1.060
3.0	2.5	0.721	−0.251	−0.797	−0.598
3.0	2.0	0.452	−0.441	−0.701	−0.283
3.0	1.0	0.070	−0.515	−0.365	−0.006
3.0	0.5	−0.026	−0.513	−0.181	0.014
3.0	0.0	−0.058	−0.510	0.0	0.0
2.0	2.0	1.000	0.0	−0.488	−1.005
2.0	1.5	0.830	−0.315	−0.517	−0.705
2.0	1.0	0.655	−0.508	−0.408	−0.408
2.0	0.5	0.538	−0.611	−0.222	−0.184
2.0	0.0	0.498	−0.644	0.0	0.0
1.0	1.0	1.000	0.0	−0.088	−0.698
1.0	0.5	0.978	−0.184	−0.076	−0.335
1.0	0.0	0.964	−0.245	0.0	0.0
0.5	0.5	1.000	0.0	−0.011	−0.353
0.5	0.0	0.996	−0.062	0.0	0.0

Table II. Representative Values

a/c	K_R	K_X
0.0	0.0	0.0
0.5	0.011	0.352
1.0	0.087	0.698
1.5	0.260	0.961
2.0	0.488	1.096
2.5	0.660	1.100
3.0	0.764	1.06
3.5	0.805	1.023
4.0	0.826	1.01
4.5	0.840	1.006
5.0	0.855	1.005
10.0	0.930	1.000
20.0	0.965	1.0
∞	1.000	1.0

Table III. Generality of Ratio a/c

Figure No.	a/c	Radius, Cm: Copper Workpiece at 68 F, 60 Cycles 450 Kc	Radius, Cm: Steel Workpiece at 2,100 F, 60 Cycles 450 Kc
2(B)	9	5.4	0.063	45.0	0.54
2(C)	3	1.8	0.021	15.0	0.18
2(D)	2	1.2	0.014	10.	0.12
2(E)	1	0.6	0.007	5.0	0.06
2(F)	0.5	0.3	0.0035	2.5	0.03

pared to reference depth. This case has been treated in detail in reference 6. A vector diagram for infinite ratio a/c, Fig. 2(A) is included for reference purposes.

Numerical Values of Coefficients

Representative values of H_r, H_t, S_r, S_t, K_r, and K_R have been calculated for various values of the ratio a/c (workpiece size in terms of reference depth) and for various values of the ratio x/c (location inside the work piece). Some of the results are tabulated in Tables I and II. They are illustrated graphically in Figs. 2(B) through (F).

Generality of Vector Diagrams

Table III illustrates the generality with which the ratio a/c describes the electrical size of a workpiece. Fig. 2(B), for example, applies equally to a copper workpiece of 5.4 cm radius at 68 F (Fahrenheit) heated by 60 cycles and to a steel load 0.54 cm in radius at 2,100 F heated by 450 kc. There are many other combinations.

As a numerical example, take the following problem:

Find the total power developed inside a 3.6-cm-diameter copper bar at 68F when it is placed inside a coil which supplies 10,000 ampere-turns per cm at 60 cycles. Find the flux density and current density at one half a reference depth below the surface. Take $\rho = 1.724 \times 10^{-4}$ ohm-cm.

Solution:

From equation 17,

$$c = 3{,}560\sqrt{1.724 \times 10^{-4}/60} = 0.6 \text{ cm}$$

$$a/c = \frac{1.8}{0.6} = 3$$

which means that Fig. 2(C) applies. From equation 16 and Table II, find the power:

$$P_2 = \frac{(10{,}000)^2(1.724 \times 10^{-4})(2\pi)(1.8)(0.764)L}{\sqrt{2}(0.60)}$$

$$= 1{,}760\ L \text{ watts}$$

or 1.76 kw/cm length.

From equation 3, Table I, and Fig. 2(C), at $x/c = 2.5$,

$$B = 4/10\pi\,(10{,}000)(0.721 - j0.251)$$
$$= 9{,}060 - j3{,}150 = 9{,}600/-19.2° \text{ gauss}$$

From equation 16, Table I, and Fig. 2(C) at $x/c = 2.5$,

$$S = 10{,}000/0.60\ (0.797 + j0.598)$$
$$= 13{,}300 + j9{,}970$$
$$= 16{,}600/36.8° \text{ amp/sq cm}$$

Magnetic Intensity

The vector diagrams illustrated in Table I and Fig. 2(A) through (F) are all based on the air-gap magnetic intensity (magnetizing force) as a reference. The diagrams drawn are to scale. The locus of the magnetic-intensity vector has a spiral shape. The magnetic intensity becomes smaller in magnitude and more lagging in time phase with increasing depth below the surface. For large diameters, the magnetic-intensity vector may make several complete revolutions between the surface and the center of a workpiece. For small diameters, the magnetic intensity at the center is hardly different, either in magnitude or phase, from that at the surface. In all cases where the ratio of radius to reference depth is finite, the magnetic intensity at the center of the work piece is larger than zero.

Current Density

In every case, the current density is maximum at the surface and zero at the center. It must be zero at the center, because it changes direction there. With increased depth below the surface, the current density vector also becomes more lagging. In electrically large workpieces, Fig. 2(A) and (B), it describes more than a complete revolution before becoming insignificant. These are the conditions under which induction heating is most efficient.

In electrically small workpieces, the current density falls rapidly to zero at the center with very little change in phase. These are more suitable conditions for transformer laminations than for induction heating.

Total Secondary Current

The optimum conditions for induction-heating a round, solid workpiece appear in Fig. 2(A), where $a/c = \infty$. The integrated total induced current I_2 has its maximum value for the applied magnetizing force. The secondary power factor is 70.7% lagging. Fig. 2(B) through (F) illustrate the progressive shrinking of this vector and the change of its phase angle as workpiece size is decreased.

Total Power Developed

The maximum power is developed in Fig. 2(A). Here the projection of the I_2 vector on the E_2 vector is largest. In spite of the fact that the E_2 and I_2 vectors come more into line with each other for smaller workpiece sizes, this is more than offset by the rapid decrease of their magnitudes.

Fig. 2(F) approaches the condition necessary for the ideal transformer lamination. However, it is well to bear in mind that hysteresis losses have been neglected in this analysis. For steel workpieces in the region covered by Fig. 2(F), hysteresis losses are probably not unimportant.

Conclusions

The ratio of workpiece radius to reference depth determines the distribution of magnetic intensity and of current density and of current density inside a workpiece being heated by induction. It is also the factor which determines whether or not a workpiece is a good induction heating load. When this ratio is small, while at the same time the workpiece permeability is high, the workpiece would serve better as a transformer lamination.

References

1. Proposed Standard, Test Code, and Recommended Practice for Induction and Dielectric-Heating Equipment. *AIEE Standard No. 54*, Oct. 1952.

2. Calculations for Coreless Induction Furnaces, **H. B. Dwight, M. M. Bagai**. *AIEE Transactions*, vol. 54, March 1935, pp. 312–15.

3. Radio Frequency Heating (book), **George H. Brown, Cyril N. Hoyler, Randolph A. Bierworth**. D. Van Nostrand Company Inc., Princeton, N. J., 1947.

4. Design of Induction-Heating Coils for Cylindrical Nonmagnetic Loads, **J. T. Vaughan, J. W. Williamson**. *AIEE Transactions*, vol. 64, Aug. 1945, pp. 587–92.

5. Design and Calculation of Induction Heating Coils, **R. M. Baker**. *Ibid.*, vol. 76, pt. II, March 1957, pp. 31–40.

6. Surface Heating by Induction, **Herbert F. Storm**. *Ibid.*, vol. 63, Oct. 1944, pp. 749–54.

Discussion

R. M. Baker (Westinghouse Electric Corporation Baltimore, Md.): Mr. Tudbury is to be complimented on the treatment of his subject. It is unfortunate that the conditions of flux and current distribution inside an induction-heated workpiece are so difficult to visualize in the variety of situations which arise in practice because it is only through such visualization that one may know what to expect and plan accordingly. This situation is undoubtably responsible for much consternation and frequent disappointment on the part of users of induction heating. The mathematics required to treat even a simple case such as the cylindrical shape chosen by Mr. Tudbury is quite involved. Yet the mathematical approach is the best way to establish quantitative concepts.

In explaining the rudiments of induction heating to the uninitiated one usually assumes that the frequency used is high enough to make the reference depth much less than the minimum dimension of the workpiece. In this case the reference depth is the effective depth of current penetration, and a simple picture of surface or skin heating results. This simple picture serves for any shape because the skin current flowing around the workpiece follows the surface contour; be it flat, round, fluted, or otherwise. One may even estimate on this basis approximate power input for a given current in the coil and foretell where the workpiece will be relatively hot or cold. It is actually a good starting point even for the expert. It represents the maximum coupling efficiency which can be achieved. Practical considerations however often dictate the use of a lower frequency and in order to anticipate results one must be able to visualize what happens as the reference depth becomes appreciable or when it may even exceed the minimum dimension of the workpiece. For certain regular shapes this can be developed on a purely mathematical basis. For more complicated shapes one may think of the actual work piece in terms of a similar regular shape which can be treated mathematically. This is, in fact, the way one gradually develops an intuitive feeling for what will happen in a particular situation even though the mathematical approach may be replaced or supplemented by the experimental heating of certain typical regular shapes in the laboratory. Everyone sooner or later learns to lean heavily on the concept of reference depth. This very useful concept is a measuring stick incorporating all the variables of the problem except the geometry of the workpiece. It permits one to translate experience from one job to another and suggests the similarity of jobs wherein the materials and frequencies used may be radically different.

All of this points to the need, especially on the part of users, for a better understanding of conditions inside an induction-heated workpiece.

It is to be hoped that Mr. Tudbury's paper will stimulate others to write on the subject.

Transient Voltages in Rectifier Transformers

B. C. BIEGA
MEMBER AIEE

H. W. LORD
FELLOW AIEE

Synopsis: This paper discusses transient-voltage phenomena which occur in rectifier circuits involving mercury-vapor hot-cathode tubes commonly used as power sources in many electronic equipments such as broadcast transmitters, high-frequency industrial-heating applications and the like. In particular, reference is made to equipments with power outputs in the range 10–200 kw with direct voltages of 3 to 15 kv.

Specifically it is intended to show the effect of transformer parameters on these transients, the voltage stresses that may occur within the transformer windings, and to suggest means for improving the capability of withstanding such transients as well as testing the ability of the transformer to withstand transients that may occur in service.

Causes of Voltage Transients in Rectifier Circuits

Faulty Operation of Tube

Sudden restriction of current flow. During normal operation, a mercury-vapor-filled tube may suddenly cease to conduct current in the forward direction either as a result of ion starvation, particularly during severe overload conditions, or when the tube is operating below normal temperature, or as a result of a pinch effect or constriction of the arc due to electromagnetic forces acting on the plasma. In any of these cases there is an extremely rapid rate of change of current which may be of the order of hundreds of amperes per μsec (microsecond). Consequently, voltages are induced in inductances such as the direct-current windings of the transformer in accordance with the equation

$$e = -L\frac{di}{dt}$$

Tube arc back. This is a spontaneous reversal of conduction in the tube. Some of the causes may be overheating of the tube, too rapid application of reverse voltage after conduction, localized cathode spots forming on the anode, application of reverse voltage of value sufficient to start a glow discharge at the anode.[1]

Frequently an arc back may start but will clear itself. However, if the arc back is sustained, this is equivalent to a short-circuit of the transformer secondary, the short-circuit current being limited only by the impedance of the supply and of the transformer (see Fig. 1). Unless these impedances are large enough to limit the short-circuit current to a safe value and unless circuit-protective devices operate rapidly, severe damage may be inflicted on the tubes, the transformer, and other circuit elements. If the load is a motor the magnitude of the short-circuit current will be substantially increased.[2]

Irrespective of the thermal and mechanical effects of the short-circuit current, damaging overvoltages may occur within the transformer windings as will be discussed later. It should be noted that an arc back may occur as a result of the high induced reverse voltages appearing at the anode following a restriction of current flow described under "Sudden restriction of current flow."

Direct-Current Interruption

In many industrial electronic circuits, particularly in high-frequency heating applications, vacuum switches are used to bias the grid of the oscillator tube in

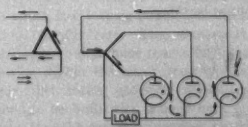


Fig. 1. Flow of arc-back current

Paper 59-149, recommended by the AIEE Electronics Committee of the Communication and Electronics Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing December 5, 1958.

B. C. Biega, formerly with the General Electric Company, is now with the Precision Transformer Corporation, Chicago, Ill., and H. W. Lord is with the General Electric Company, Schenectady, N. Y.

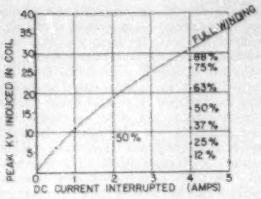


Fig. 2. Induced voltages in coil due to sudden interruption of direct current. Percentages indicate voltages at tap in winding

order to obtain a sharp cutoff to the power tube. Occasionally this may be done as a protective measure, but in other cases this method is used as a means of controlling the heating cycle. The very high rate of change of direct current induces an overvoltage in any inductances in the circuit including the transformer d-c windings.

Fig. 2 illustrates the relationship of peak-induced terminal voltages to the magnitude of the direct current suddenly interrupted in a coil of a 60-kva transformer rated 5,150 volts line-to-neutral. To obtain these data, direct current from a low-voltage external source was circulated through the winding and interrupted with an *RG9* vacuum switch. The voltage rise time was of the order of 100 μsecs.

Alternating-Current Interruption

In many industrial applications involving repetitive cycles of power application, the duration of the cycle is controlled by contactors in the primary lines of the transformer. When the contactor is opened the load current normally continues to flow through the arc which is maintained between the parting contacts until the current passes through zero. Occasionally, and in particular on light loads, the current will extinguish suddenly before current zero is reached.

The electromagnetic energy stored in the magnetic field of the primary winding, $W = {}^1/_2 L_{l,p} i^2$, is transferred to electrostatic energy by charging the capacitance of the transformer primary winding. The maximum value of the induced voltage becomes

$$e = i\sqrt{\frac{L_{l,p}}{C_p}}$$

where

$$\sqrt{\frac{L_{l,p}}{C_p}}$$

is the surge impedance of the transformer. In these equations

i = instantaneous value of the current at the moment of interruption
$L_{l,p}$ = leakage inductance referred to primary
C_p = distributed capacitance of primary winding

Inrush Current

When the primary contactor is closed the initial inrush of the transformer magnetizing current may attain a momentary peak value up to 25 times rated full-load peak current. The inrush current is very highly damped and declines to one-tenth of its initial value in a few cycles. The peak value of the inrush current is a function of the value of the residual magnetization of the transformer core, the normal operating peak flux density and the air-core inductance of the primary coil. In addition it depends upon the instantaneous value of the line voltage at the instant at which the contactor closes and also upon the impedance of the supply line.[3] The high electromagnetic forces resulting from the peak inrush current may force the contacts apart. In this case a sudden interruption of the alternating current may then result with the effects described under "Alternating - Current Interruption." High voltages may be induced not only in the a-c windings of the transformer but also in the d-c windings. Fig. 3 shows transient voltages induced in the output of a 12-kv full-wave rectifier due to transformer inrush current.

Retarded Firing of Rectifier Due to Grid Control

In full wave and multiphase rectifier circuits, each tube fires in turn as its anode voltage becomes more positive than that of the preceding tube. The current flow is transferred from one phase to the other at this time. The rates of decay and rise of current in each phase are controlled by the commutating reactance of the transformer. During this period both the rectifier tube which has been carrying current and the one that is beginning are conducting current simultaneously, consequently the transformer windings are short-circuited. This means that the ends of the windings are held at some voltage between the normal induced values. At the end of commutation this short circuit is removed and the winding-terminal voltages then suddenly assume their proper values. The rate at which the winding that has ceased to carry current assumes its normal voltage is determined by the

Fig. 3. Transients in rectifier output voltage due to transformer inrush current. Vertical scale, 12 kv per division

leakage inductance and distributed capacitance of that winding. If the L-C circuit formed by this winding has a relatively high Q the voltage will overshoot by a value that nearly equals the difference voltage that existed at the instant current ceased to flow. In many rectifier applications grid-control tubes are used to reduce the output voltage. As the firing angle of such a tube is retarded these voltage differences become increased in magnitude. The voltage transients repeat each cycle and, therefore, produce a continual overstressing of the transformer winding, or may even cause arc backs in the rectifier tubes.[4]

Effect of Transformer Reactance on Transient Conditions

The leakage reactance of the transformer windings is the main factor controlling the magnitude of fault currents due to a short circuit within the d-c circuit or due to a rectifier-tube arc back. These fault currents, if allowed to exceed reasonable values, will destroy the tubes and also cause damage to the transformer windings themselves as a result of severe overheating and particularly as a result of the enormous electromagnetic forces acting upon the windings.

From this point of view the leakage reactance should be high, preferably about 10%. Such a high value of reactance, however, would result in poor regulation and from this point of view is undesirable.

Likewise, high reactance is undesirable since it increases the peak voltages in-

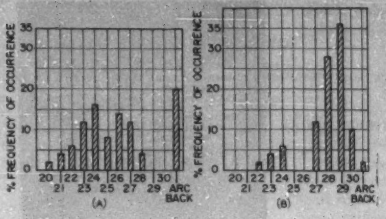


Fig. 4. Transformer terminal-to-ground voltages due to interruption of 3.5 amperes d-c in output of 12-kv 6-phase double-way rectifier. Peak kv induced in one phase

A—Transformer impedance, 3.1%

B—Transformer impedance, 4%

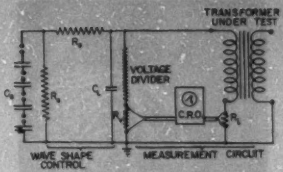


Fig. 6. Diagram of Marx circuit-impulse-test set

duced at the secondary terminals as a result of current interruption in the d-c circuit or at the primary terminals due to current interruption at the primary contactors. The secondary overvoltages appear at the anode as excessive reverse voltages and may produce an arc back of the rectifier tubes.

A vacuum switch was inserted in series with the load supplied by a 6-phase double-way rectifier. This vacuum switch was opened a large number of times and the peak induced voltage between one of the secondary transformer terminals and ground was recorded. Fig. 4 shows the relative number of occurrences of various values of peak voltages for two different values of transformer reactance. The higher value of reactance was obtained by means of air-core inductance coils inserted in the primary leads of the transformer. The number of arc backs, resulting in opening of the primary circuit breaker, is also recorded for each case. Overvoltages undoubtedly occurred every time the vacuum switch was opened but they did not always occur in the phase to which the oscilloscope was connected. From visual observation, it was obvious that incipient arc backs occurred with virtually every operation of the vacuum switch. The fact that complete flashover resulting in opening of the primary circuit breaker occurred less frequently with the higher reactance, in spite of the greater frequency of higher overvoltages, may be attributed to the lower value of fault current.

The selected value of transformer reactance must, therefore, be a compromise between the requirements of low regulation and limitation of fault current. The type of circuit used and the type of operation also should be considered. In applications where frequent switching with primary contactors is contemplated, the use of primary inductance coils for controlling the over-all value of transformer reactance should be considered, as these coils help to limit the peak values of primary inrush current.

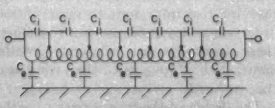


Fig. 5. Winding capacitances

C_i = internal distributed capacitance

C_e = external distributed capacitance

Impulse Waves and Voltage Distribution Within the Transformer Windings

Winding Capacitance

A transformer winding, in addition to inductance, also possesses capacitance. Fig. 5 shows a simplified representation of the winding capacitances. While they are shown here as lumped values, in actual fact they are, of course, composed of an infinite number of individual capacitances from each turn to ground (the external distributed capacitance) and from each turn to every other turn (the internal distributed capacitance). In a winding devoid of capacitance, voltage transients travel through the winding due to the inductive coupling between turns, resulting in uniform voltage distribution throughout the winding. In actual fact, such voltage transients travel through the winding at a speed dependent upon the relative values of the various capacitances. In addition, because of inductance, the waveshape of the transient is modified as it travels through the winding and oscillation will occur.[5,6] Thus it is not sufficient to know only the magnitude of the voltage transients that may occur in a given rectifier circuit in order to design a satisfactory transformer. It is also necessary to know in what way these transients will be distributed through the winding, that is, to what maximum value of voltage each individual piece of insulation (turn insulation, layer insulation, section insulation) might be subjected during these voltage transients. Both the steepness of the wavefront and the physical winding configuration affect this voltage distribution.

Impulse Testing

In the past 40 years, considerable theoretical and experimental work has been done to determine the effect of winding configuration and design upon voltage distribution. In this field of engineering the predominant cause of concern has been the abnormal voltage transients arising when lightning strikes the transmission line to which the transformer is connected. It has been determined that an impulse wave which rises to peak value in $1^1/_2$ μsecs and decays to half value in 40 μsecs is representative of the most severe lightning-surge conditions that may arise in service. This type of transient can be duplicated in the laboratory by means of the Marx circuit-impulse generator shown schematically in Fig. 6.

The operation of this circuit is briefly as follows:

The capacitors C_G are charged from a rectifier to the desired value while connected in parallel. They are then switched in series and discharged into the circuit. The voltage across the load (the transformer winding being tested) rises until the potentials across C_G and C_L are equal. This portion of the impulse is the wavefront. The steepness of the wavefront is controlled by the series resistance R_S. C_G and C_L are now in parallel and discharge through the load resistance and R_G. The rate of decay (the wavetail) is varied by adjusting R_G. Since the constants of the transformer winding also affect the impulse waveshape R_S and R_G must be adjusted for each new transformer design.

To determine the required values of R_S and R_G, a transient analyzer is used. This is a scale model of the Marx circuit in which low-voltage pulses are repeated at a 60-cycle rate. This equipment is also used to determine the voltage distribution within the transformer windings. For this latter purpose special coils need to be built with a large number of taps brought out from various parts of the winding.

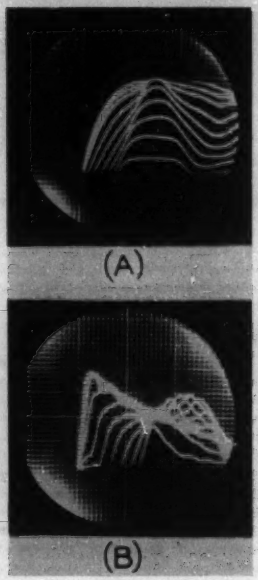


Fig. 7. Time-voltage relationships in coil

A—10x100-μsec wave impressed at terminal. Horizontal scale 1 μsec per division

B—0.6x10-μsec wave impressed at terminal. Horizontal scale 0.5 μsec per division

It has been considered that the same approach used in power transformers also be used for determining the adequacy of design of rectifier transformers. In power transformers a basic impulse insulation level (BIL) is used to determine the adequacy of the design and insulation system. Thus a transformer with a BIL of 90 kv has been tested and shown to be capable of withstanding the standard $1\frac{1}{2} \times 40$-μsec inpulse waves of a crest value of 90 kv. However, cases have occurred where rectifier transformers with d-c windings of 50-kv BIL and even 90-kv BIL have failed in rectifier circuits with a 12-kv direct-voltage output. This approach by itself is therefore demonstrated to be inadequate.

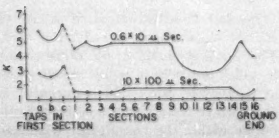


Fig. 8. Distribution of K factor within winding. As shown by taps in first section, stresses within each section may be higher than indicated

Effect of Waveshape on Voltage Distribution

The steepness of the wavefront has considerable effect upon the voltage distribution within a coil. If we assume a constant applied voltage and a constant velocity of travel of voltage through the coil due to the combined capacitive and inductive coupling between portions of the winding, the faster the rate of rise in potential at a given point and the greater will be the ratio that exists between this rate of rise and the rate of propagation. Thus, the steeper the voltage wavefront, the higher the instantaneous potential difference that arises between two portions of the winding. Figs. 7(A) and (B) illustrate this point dramatically. In these oscillograms the potentials to ground at a number of taps within a pie-wound coil due to an impulse wave impressed on one end of the winding are superimposed. In Fig. 7(A) the shape of the impulse wave is 10×100 μsecs, in Fig. 7(B) it is 0.6×10 μsecs. In each case the crest voltage is identical. The instantaneous potential existing between any two taps can be obtained by subtracting the corresponding instantaneous values from one another. An arithmetical representation of the nonuniformity of voltage may be obtained by calculating the voltage distribution factor

$$K = \frac{\text{Instantaneous volts between taps as per cent of impressed crest value}}{\text{Turns between taps as per cent of total winding}}$$

for the condition of maximum instantaneous potential appearing between any two taps. (In a low-frequency induced test $K = 1$.)

Values of K are plotted in Fig. 8 for two different impulse waves impressed on the same pie-wound coil. From this graph it is obvious that a coil of this design might pass a test in which the impressed impulse wave has a slow front, but would fail when tested with an inpulse wave of a much steeper wavefront, as portions of the insulation are so much more severely stressed by the latter test.

Winding-Voltage Stresses in a Rectifier Circuit

Induced Voltages in the Transformer Windings

Voltage transients due to sudden changes of current are induced within the windings of the transformer (or inductors) in accordance with the equation

$$e = -L_{l,s}\frac{di}{dt}$$

Due to capacitive effects oscillations may arise of frequency

$$f = \frac{1}{2\pi\sqrt{L_{l,s}C_s}}$$

where

$L_{l,s}$ = leakage inductance referred to secondary

C_s = distributed capacitance of secondary winding

The values of $L_{l,s}$ and C_s are such that these transients are relatively slow, with rates of rise from zero to crest of the order of at least 50 μsecs. Therefore, these transients are relatively evenly distributed throughout the windings, that is, $K \approx 1$.

The design of the transformer to withstand these induced-voltage transients is a relatively simple matter provided the maximum possible value of the induced

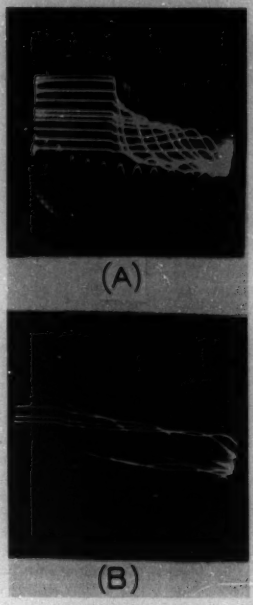


Fig. 9. Time-voltage relationships within winding during arc back

A—Horizontal scale 0.3 μsec per division

B—Horizontal scale 0.1 μsec per division

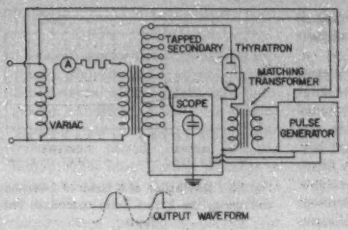


Fig. 10 (left). Circuit used to simulate arc back

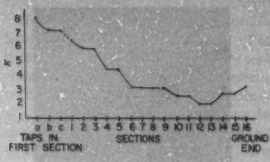


Fig. 11 (right). Distribution of K factor in 16-section coil during arc back

voltage can be predicted. For some values of interrupted direct current, peak-induced voltages at taps within the coil are also shown in Fig. 2. Since the voltage distribution is essentially uniform the voltage stresses turn-to-turn, layer-to-layer and between sections can be calculated and the required insulation provided. Voltage stresses between layers are best reduced by splitting the coil into a large number of sections or "pies," thus providing a large number of layers each with a small number of turns. The insulation, or spacing between sections, must be sufficient to withstand the voltage gradients between them.

Voltage Stresses Due to Arc Back

When the tube arcs back, the anode potential suddenly becomes equal to the cathode potential (neglecting tube drop). The potential existing across the coil likewise collapses. The rate of voltage collapse at the secondary line terminal may be of the order 0.1 to 0.2 μsec. From the point of view of voltage distribution within the coil this is equivalent to impressing upon the coil an impulse wave of 0.1- to 0.2-μsec front and a very long tail.

Figs. 9(A) and (B) show the change of potential at a secondary terminal and at the various taps of a coil under simulated arc-back conditions. Fig. 9(B) is identical to 9(A) except that the sweep rate is increased to emphasize the potential differences existing at the line end of the coil. The induced voltages at each of the taps appear as straight lines because of the speed of sweep compared to the 60-cycle induced voltage. The coil is the same as that used to obtain Fig. 7. The diagram of connections for this test is shown in Fig. 10. Briefly, the method used is as follows: The transformer primary is excited continuously from a 60-cycle source. The thyratron is triggered to fire at the peak of the induced voltage in the transformer secondary. The sweep of the oscilloscope is initiated slightly earlier. A resistance in the primary circuit is used to limit the short-circuit current. By means of this test circuit the arc back of a tube in a 3-phase single-way Y circuit may be duplicated.

It will be seen from Fig. 9(A) that the steepness of the wavefront is quite rapidly attenuated, but because of the low internal distributed capacitance of the pie-wound winding, several μsecs elapse before the voltage collapses in the last of the pies. The maximum voltage appearing across the first section is equal to about 36% of the total peak induced voltage across the whole coil. Fig. 11 shows the K values for this test which may be compared with Fig. 8.

In a 6-phase double-way circuit two tubes carry load current in series when operating normally. When an arc back occurs, arc-back current from the transformer must flow through the tube of the other phase which is normally conducting, thus the collapse of voltage occurs across two terminal-to-neutral phases in series. To simulate this condition a 3-phase circuit similar to that of Fig. 10 was set up. In this case the thyratron was connected between the two line terminals of two phases. The voltage-versus-time relationships at the taps is shown in Fig. 12. Voltages for all of the taps are not shown. However, inspection reveals that the voltage distribution within the coil which is connected to the anode of the arcing tube is much the same as in the previous case. The other phase is protected by the first phase since the steepness of wavefront (voltage collapse) has been considerably reduced by the time the neutral point is reached.

From the point of view of internal voltage stresses within the coil, it is seen that the most serious condition occurs during tube arc back. In turn, arc back is most likely to occur when an inverse voltage in excess of its rating is impressed upon the tube such as by an excessively high voltage being induced at the transformer terminals due to a sudden change of direct current or some other cause. Then, if the normally conducting tube, which must then carry a heavy short-circuit current, should suddenly cease to conduct, further and still higher voltages may be induced in the transformer coils and cause even greater damage to their insulation system. During the tests that were run to obtain the data of Fig. 4, simultaneous arc backs on as many as three tubes were observed.

Maximum Expected Voltages

The peak value of the terminal voltage induced in a coil has been shown to be a function of the rate of change of current and also the inductance and surge impedance of the windings. If all these factors are known the peak induced voltage can be calculated. In the higher power rectifiers the calculated values can easily reach 100 kv or more under the worst predictable conditions. At such high voltages some circuit component invariably breaks down or at least a flashover occurs at insulators or exposed terminals. Also, if the peak voltage is not limited by some device such as nonlinear resistors, lightning arresters, spark gaps, or capacitor-resistance networks, such induced voltages will generally exceed the inverse-voltage rating of the tube and cause an arc back. Therefore, in the absence of any other voltage-limiting devices, the actual arc-back voltage of the tube is the maximum peak voltage which must be considered in transformer design.

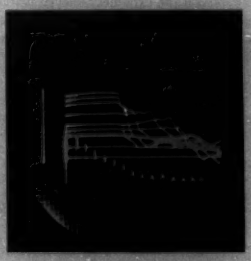


Fig. 12. Time-voltage relationships with arc back across two phases. Horizontal scale, 0.5 μsec per division

Fig. 13. Time-voltage relationships in 2-section coil during arc back. Turns spaced in first layer. Horizontal scale, 0.1 μsec per division

Unfortunately, tubes themselves are quite erratic and cannot be depended upon to arc back at any specific voltage level since this depends upon the age of the tube, the load current, and the temperature of the tube. In general, however, most high-voltage rectifier tubes should be expected to arc back at a voltage which does not exceed about twice their inverse-voltage rating. Transformer coils designed to handle induced voltages with a peak value equal to twice the inverse-voltage rating of the tube and to successfully survive the internal voltage stresses (and also mechanical stresses) resulting from arc back at such a voltage level should certainly outlast several sets of tubes.

Preventative Measures

Coil Design

The coils must be designed to withstand the previously discussed voltage stresses. Their capability to do so will be much enhanced if the extremely uneven voltage distribution shown to exist, for example, in the pie-wound coil with a large number of sections, can be improved.

The initial voltage distribution is a function of

$$\alpha = \sqrt{\frac{C_e}{C_i}}$$

where

C_e = external distributed capacitance
C_i = internal distributed capacitance. See Fig. 5

The lower the value of α, the more the voltage distribution within the coil approaches uniformity for any given waveshape. When $\alpha \approx 0$ the voltage gradient across the whole winding is essentially uniform, or $K \approx 1$ in all parts of with winding.[5,6]

From this it is seen that the internal distributed capacitance should be large compared with the external distributed capacitance. This condition is approached in a simple layer-wound coil. Fig. 13 shows the relative voltage distribution in a coil wound in two sections. This coil has the same physical dimensions and exactly the same number of turns as the 16-section coil used to obtain Figs. 8 and 11. In this case taps were brought out at the ends of the layers. Comparison between the figures reveals that the voltage distribution is, as expected, much more uniform. Fig. 14 shows the values of K throughout the winding.

However, this type of winding is not necessarily the best suited to withstand rectifier transients. Although the voltage distribution of the 2-section type is much more uniform for a given coil geometry, the actual layer-to-layer voltage stresses are higher, both under the induced condition and the arc-back condition, than in the multisection coil. An optimum design is obtained by adjusting the total number of sections so as to provide a reasonable compromise between transient-voltage distribution and normal interlayer voltage stresses.

Another approach to the voltage distribution problem is to use electrostatic shielding. External shields connected to the terminal lead in effect neutralize the external distributed capacitance of the winding and thus reduce α. Theoretically, it is possible to design a series of such shields spaced in such a manner as to obtain uniform transient-voltage distribution, but it is not practical to do so in transformers of the size under consideration.

An approach directly opposite to that of electrostatic shielding is to space the first few turns of the winding adjacent to the terminal end. This in effect decreases the capacitance which results in increased voltage stresses between spaced turns. This develops into a race between added insulation and increased voltage stresses. However, this method does provide some benefits as it reduced the steepness of the wavefront applied to other parts of the coil. This effect is seen in Fig. 14 in which the turns in the first layer were spaced apart.

Protective Devices

Since the severest voltage stresses in the coils are encountered as a result of arc back, the likelihood of this phenomenon occurring should be reduced. If induced voltages can be prevented from exceeding the rated inverse voltage rating of the tube, the probability of an arc back occurring can be substantially reduced. This improvement is also desirable from the point of view of tube life, and over-all performance of the circuit.

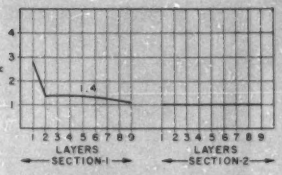


Fig. 14. Distribution of K factor in 2-section coil during arc back. Turns spaced in first layer

Spark gaps set to flashover at a predetermined value will achieve this if series resistors of proper values are used to limit the current in the arc. Such spark gaps should be sealed to prevent tampering and to reduce variations due to atmospheric conditions and surface contamination.

Nonlinear resistors such as Thyrite* are very effective in limiting overvoltage. However, in higher voltage circuits the number of such nonlinear resistors that need to be used in series becomes prohibitive in cost and space requirements.

In the higher voltage circuits, a combination of nonlinear resistors and sealed sparkgaps offer a satisfactory solution at reasonable cost. Such combinations are available in the form of standard lightning arresters used to protect distribution transformers from lightning and switching surges, for example Magnevalve* arresters.

For the higher voltage rectifiers, R-C networks also offer a solution. However, they should be individually designed for each application and circuit to obtain the desired characteristics.

All such protective devices should be connected across each winding of the transformer with leads which are as short as possible. In delta-connected transformers and even Y-connected transformers feeding 6-phase double-way rectifiers, neither end of the transformer winding is grounded. This requires some modification of the mounting of standard lightning arresters since they are designed for normal operation with the lower terminal grounded.

* Registered trade-mark of the General Electric Company.

Testing

Present standard test procedures provide for adequate testing of major insulation between windings and winding-to-ground. However, none of the existing standard test procedures for rectifier transformers provide for an adequate testing of the direct current windings to withstand the kind of voltage transients discussed in this paper. Standard induced-voltage procedures provide for testing at twice normal rated voltage. Unless protective devices are provided that will definitely limit induced voltages to such a value, it is evident from the preceding discussion that this value is much too low. Furthermore, a 400-cycle induced test gives no indication of the ability of the winding to withstand steep-front or high-frequency transients of even this low a value.

The authors propose that an adequate test procedure should consist of inducing each direct current winding to a peak value equal to two times the rated-peak inverse-voltage rating of the tubes to be used in the circuit (the purchaser would specify this value in each case). In the event that protective equipment is to be provided, the peak value of the induced-voltage test could be reduced to a value equal to 120% of the overvoltage cut-off point of the protective device.

Arc backs at this voltage level would be simulated by a circuit similar to that shown in Fig. 10 except that cascaded hydrogen Thyratrons would replace the one shown. The Thyratrons would be triggered to fire at the peak of the induced voltage wave at least once every second. The induced voltage test would need to be at a frequency of 400 cycles per second and should be continued for 10 seconds. Failure of the windings might be detected by observing the input current on the primary side of the transformer.

An alternative method might consist of an induced test of the same value as described before, followed by a chopped-wave impulse test of crest value equal to 110% of the peak of the induced voltage, followed by a further induced-voltage test. The impulse wave would be impressed on the high-voltage (direct-current) windings of the rectifier transformer.

Conclusions

This paper has attempted to bring into proper perspective the relationship between the transformer design and the other elements of a power-rectifier circuit in which mercury-vapor-filled tubes are used as the rectifying element. To obtain satisfactory over-all performance, co-ordination is required between the manufacturer of the rectifier tubes, the transformer, and of the ultimate equipment. It is believed that many of the difficulties that arise from time to time could be avoided if the designers of each of these components and systems were better acquainted with the characteristics and requirements of each other's devices.

All too frequently, specifications for a rectifier transformer spell out nothing more than the input and output voltages and the output current ratings. From this meager information, the transformer designer is expected to design a satisfactory transformer. However, transformers of identical designs sometimes perform satisfactorily in one application but give trouble in another.

No rigorous mathematical solutions have been provided in this paper. The data given are based on actual tests and should be considered only as reference points indicating the order of magnitude of some of the phenomena that may be encountered. Many of the references go into much greater detail of particular aspects of the problem. The authors feel that much more work needs to be done in this area and specific data should be gathered, both in the field and in the laboratory, to assist the designers both of the transformers and over-all power-rectifier circuits in obtaining optimum designs.

Finally, specific test methods and values need to be established in order to give the transformer designer a specification value to use for design purposes and a method of checking the product. This will also assure the equipment manufacturer that he is obtaining the required quality. It is hoped that the procedures and values suggested here may serve as a useful starting point for the establishment of such standards.

References

1. Industrial Electronics Reference Book (book). John Wiley & Sons, Inc., New York, N. Y., 1948, p. 52 ff.

2. Rectifier Fault Currents—II. C. C. Herskind, A. Schmidt, Jr., C. E. Rettig. *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 243–52.

3. Electronic Transformers and Circuits (book), Reuben Lee. John Wiley & Sons, Inc., 1947, p. 100.

4. Voltage Impulses in Rectifiers, T. Douma. *Philips Technical Review*, Eindhoven, Netherlands, vol 9, no. 5, 1945, pp. 135–146.

5. The Insulation of Electrical Equipment (book), Willis Jackson. John Wiley & Sons, Inc., 1954, pp. 259–77.

6. Transformer Engineering (book), L. F. Blume, A. Boyajian. John Wiley & Sons, Inc., pp. 463–81.

Discussion

Reuben Lee (Westinghouse Electric Corporation, Baltimore, Md.): The authors have outlined several of the transient-voltage conditions in rectifier transformers, some of them rather exhaustively. With most of the information in this paper those familiar with high-voltage rectifier design will agree. A few points appear to deserve further attention.

1. Arc back is treated as an abnormal condition "likely to occur when an inverse voltage in excess of its rating is impressed upon the tube." It is further stated that "if induced voltage can be prevented from exceeding the rated inverse voltage of the tube, the probability of an arc back can be substantially reduced."

Prevention of arc backs is part of rectifier design and the inverse-voltage rating of the tube should take into account the possibility of arc back in gas-filled rectifiers. Similarly, the transformer should be so insulated that it will be the least likely circuit element to fail in the event of arc back or any voltage transient.

2. For "transformer coils designed to handle induced voltages with a peak value equal to twice the inverse-voltage rating of the tube," an induced voltage of twice normal value is recommended. But the basis for this part of the recommended test procedure is uncertain. Under average service conditions it may be adequate if backed up by some sort of short-duration test, but there may be applications such as laboratory rectifiers where this induced voltage would be unduly severe. It would be interesting to know why a general twice-normal induced voltage test is proposed.

3. The example given concerning BIL seems hardly to demonstrate the inadequacy of BIL for rectifier transformers. For 12-kv direct-voltage output, 15-kv insulation class would be used by experienced designers. In an oil-immersed transformer below 500 kva the American Standard Power Transformer Standard[1] specifies 110-kv chopped wave or 95-kv full-wave impulse test. If the transformer in question had been designed according to this standard it might not have failed.

4. Before adopting the test similar to Fig. 10, in which the surge has shorter duration than standard impulse tests, additional testing would be necessary to correlate results with known levels of insulation. Furthermore, the thyratron simulates gas-filled rectifier operation, and would not apply to high-vacuum and solid-state rectifiers. For these latter types, either the present standard BILs should be used or still other standards formulated.

Reference

1. Transformers, Regulators, and Reactors. *American Standard C57.12*, American Standards Association, New York, N. Y., 1958.

Andrew A. Halacsy (Federal Pacific Electric Company, Newark, N. J.): For better clarity I recommend the use of e_p and e_t in the equations giving the value of the induced voltage. In the usual manner, the authors stressed that the capacitance of the winding has to be considered as distributed among the parts of the winding. It would be interesting to investigate the effect of the nonuniform distribution of the leakage inductance, which is the case when the spacing of the turns within the same coil is not uniform or when the position of some primary turns relative to the secondary coil differs from that of other turns, such as those at the ends and in the axial middle of a coil.

The circuit shown in Fig. 10 is recommended to simulate an arc back. It does so except the direction of the reversed current. This is clear when Figs. 1 and 10 are compared. A reversed current, of course, will create the same voltage transient but of reversed polarity. Breakdown of insulation depends on the polarity of such steep voltage waves. To restore the simulated conditions a reversal of the connections of the secondary coil is required.

In the paragraph under "Inrush Current" it is stated that "the peak inrush current may force the contacts apart." How often was this fault of the circuit breaker observed and in what kind of breaker?

B. C. Biega and **H. W. Lord**: Mr. Lee in his first point emphasizes once more the importance of proper co-ordination between the design of the transformer, the tubes, and the over-all equipment.

The induced-voltage test of peak value equal to twice rated inverse voltage of the tube is recommended on the premise that, if no other voltage-limiting devices are included in the circuit, induced voltages of this magnitude are likely to occur before the tube arcs back. Induced voltages of such magnitude may be caused by any of the circumstances described in the paper.

The 50-kv BIL referred to in the paper is that recommended by American Standard for dry-type transformers of the 15-kv insulation class. But a 95-kv or even 110-kv standard-impulse test does not guarantee that a transformer will stand up under arc-back conditions in rectifier service where the wavefront is so much more severe.

With regard to item 4 of Mr. Lee's discussion, the proposed test procedures are intended for transformers working with mercury-vapor rectifier tubes. It is agreed that much more work needs to be done.

Dr. Halacsy's point on polarity is correct, but in this particular case relative potential with respect to the ground plane is of lesser importance than the internal momentary voltage distribution within the coil.

The forcing apart of contacts by the peak inrush current has not been physically observed. However, the transient voltages observed can hardly be explained if this is not assumed. This occurred with standard contactors used for control of a particular circuit in which the transients shown in Fig. 3 were observed.

Significance of Heat Pump Coefficient of Performance

CONSTANTINE W. BARY
FELLOW AIEE

Synopsis: This paper deals with the load aspects and economic aspects of three types of residential heat pump systems which employ supplementary heat sources. It contains illustrations of their load characteristics over a wide range of annual COP (coefficient of performance), depicts relationships of the relative costs to serve these loads from long-range and short-range considerations, and provides, as a reference, load characteristics and economic data for three comparable types of resistance heating systems with their complementary electrical cooling loads, and a band of competitive operating costs to customers using gas for heating and cooling.

THE COP of a refrigerating system is defined as the ratio of the heat absorbed at low temperature to the heat equivalent of work done. For an electric heat pump, operating on the heating cycle, it may be stated as the ratio of the total useful heat output (heat absorbed at low temperature plus the heat equivalent of work done) to the equivalent heat of electric energy input. The COP's general thermodynamic principles are well known and its theoretical and practical aspects have been covered in the technical literature.[1] Throughout this paper, and to simplify the presentation, the abbreviation COP will be used for this ratio as applying to conditions of an electric heat pump on its heating cycle, including its auxiliaries, but excluding supplementary heat sources, such as resistors or heat storage means. When prefaced by the word "annual" the term means the integrated results over a period of one year.

Over the past 10 years, the application of the heat pump to provide year-round comfort in the home has grown gradually in the United States. Much has been written on the subject including a number of papers by the author on its significance as a residential electric load, and the economic aspects thereof.[1-4] But continued research, in the light of almost universal practice among manufacturers of providing relatively large amounts of supplementary resistance heating in conjunction with air-source heat pumps, has revealed some new load and economic aspects which are believed to be of vital significance in the economics of supplying electricity to this device. These must be fully realized by all concerned if the electrical heat pump is to take its economic place as a mass-market device in providing comfort through the heating and cooling of homes, and thus fulfill the aspirations of many thoughtful men in electric utility industry for all-electric home living.

The object of this paper is to set forth the results of research on the load aspects and economic aspects of several residential heat pump systems which employ supplementary heat sources; and to show a relative comparison of the economics of the utility's cost of service supply over a wide range of annual COP. (Comparison of first cost of equipment and its maintenance expense are not included because they represent different facets of the problem and do not affect the features related in this paper.)

All data used in the paper reflect either the results of actual experience, or the careful research which the author, his colleagues,

Paper 59-20, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted September 18, 1958; made available for printing November 13, 1958.

CONSTANTINE W. BARY is with the Philadelphia Electric Company, Philadelphia, Pa.

The author acknowledges the valuable assistance and suggestions of A. Abramovitz, W. C. Astley, and J. A. Thielman, all of the Philadelphia Electric Company.

and others in the industry have conducted over many years. They reflect the fundamental concepts and well-tested techniques which have been established and used in studying the behavior of electric loads, and the costs involved in serving them.[5-7] Actual experience has proved them to be reliable guides in determining the economics of the service rendered by systems with which the author is familiar. Although the basic data used herein apply to conditions of a large metropolitan system operating in the northeastern part of the United States, and the results obtained for that particular system will undoubtedly differ in detail from those that may be derived for other systems, nevertheless, it is believed that the general character of the results contained herein is of fundamental nature and should be useful to others.

Load Characteristics

The load characteristics used in this paper have been derived painstakingly, as previously described[4] for the same size house of 55,000-Btu-per-hour heat loss as that described in reference 4, and for the same significant features of climatological characteristics. These load characteristics data are summarized in Table I for three types of heat pump systems, and for three types of resistance heating systems with their complementary electric cooling loads. Each type of heat pump employs air as a heat source or heat sink, and each is sized at 60% of design heating requirements, which is ample to meet the design cooling requirements of the climatological conditions of the area. Types *A* and *B* employ resistors as supplementary heat sources to satisfy the peak heating requirements, and type *C* employs a heat storage system for that purpose.[8] Resistors of type *A* are sized for, and the system operates under, the conventional night set-back of indoor thermostats with quick heat recovery in the morning, as now commonly practiced in homes with existing fuel-fired systems. The resistors of type *B* are sized for, and the system operates under, a fixed setting of the indoor thermostat which, perhaps, residential customers could be induced to follow with a heat pump installation. Type *C* employs a heat storage system which embodies resistors for charging. These resistors are electrically interlocked with the heat pump so as to permit their operation only during the time when the heat pump is not operating.[8,9] For purposes of this study it is assumed that changes in the annual COP of the heat pump would be reflected in the same relative degree over the entire temperature range of heat pump operation.

To provide a reference point for the load characteristics of heat pumps, Table I also contains data on the load characteristics of three comparable types of resistance heating systems with their complementary electric cooling loads. Type *D* is sized for, and operates under, the conventional night set-back of indoor thermostat with a quick heat recovery in the morning; type *E* under fixed setting of thermostat. Type *F* is sized for 60% of the design heating requirements, as is the case with the type *C* heat pump, and it employs a heat storage system to satisfy the peak heating requirements. Such storage system embodies resistors for charging, which are interlocked with the main resistance heaters so as to permit their operation only during the time when the main heaters are not in use.

Three levels of diversified demands are shown for each of the types of installations: 1. the maximum, which determines the economics of long-range considerations, and which occurs in the morning on an average day of the coldest spell in the winter; and 2. two smaller values, which determine the economics of short-range considerations and occur either in the evening of such a day or in the evening of an average day of the warmest spell in the summer.

The following important features are revealed by these data:

1. The diversified annual load factors of types *A* and *B* systems, decrease with an improvement in the annual COP but, for type *C*, the annual load factor increases with such improvement.

2. The diversified annual load factors of systems which employ heat storage are much higher than those of other systems shown.

3. The magnitudes of maximum demands and energy use of all types of heat pump systems are substantially lower than those of comparable types of resistance systems; and the spread increases with an improvement in the annual COP.

4. The coldest-spell evening diversified demands of all types of heat pumps are at the level of the maximum diversified demand of the heat pump with supplementary heat storage (type *C*). The warmest-spell evening diversified demands of all types of heat pumps are all at the same level, and of a magnitude which is about 60% of the maximum diversified demand of type *C*.

5. Under present-day levels of annual COP, the energy use by supplementary resistors is relatively small, but its kilowatt demand is high. With increasing annual COP the significance of these quantities increases. Although the energy use by supplementary heat storage is larger than by supplementary resistors, contribution to the demand of the system is zero.

6. The coldest-spell evening diversified demands of resistance type systems are substantially higher than those of corresponding heat pump systems. But their warmest-spell evening diversified demands are generally in line with the demands of heat pump systems for annual COPs in the order of 3.

Long-Range Economic Considerations

It is a well-established principle in the economics of utility service that the use of the maximum diversified demand of a large-sized electric load application, such as house heating, determines the fundamental nature of the utility's cost to serve such load, independent of any temporary effects that may be created by, or on, the load characteristics of the class of service of which the load is a part. This principle reflects thus the economics of service supply under long-range considerations. Based on load characteristics shown in Table I, and employing the maximum diversified demands, it is possible to depict the trends in the utility's added annual cost to serve the load of each of the three types of heat pump systems, with improvement in their annual COP, and of each of the three types of resistance systems. These cost trends, computed from the full-cost-to-serve relationship shown in Fig. 1 and described in reference 4, are shown in Fig. 2(A) and (B). They are computed under the assumption that the 100% datum point in Fig. 1. is at a level of two cost units per kw-hr (kilowatt-hour), the basis for which was also described in reference 4.

Total Cost

Fig. 2(A) depicts the trends in total annual costs. For a better understanding of their significance it also provides a reference band of competitive operating costs to customers, which was computed for a home with the same heating requirements as stated in this paper, using gas for heating and cooling, including electricity for auxiliaries and water, all at existing rates in the area of the system described herein.

The following important features are revealed from such long-range considerations:

1. Improvement in the annual COP of heat pumps has a pronounced effect upon making the cost to serve the load of each of the types analyzed come closer to the competitive reference band. However, only the heat pump system with supplementary heat storage approaches this reference band in the region of annual COP of 3, and drops below that band beyond that region.

2. Heat pump systems employing supple-

Table I. Illustrative Characteristics of Electric Load of Heating and Cooling an Average-Size Home

	Heat Pump Annual COP for Heating	Annual Noncoincident Maximum Demand, Kw			Annual Diversified Demand, Kw			Annual Energy Use, Kw-Hr × 10^3			
		Total	Heat Pump Portion	Resistors Portion	Maximum 1	Winter, 6 P.M., EST* 1	Summer, 6 P.M., EDT† 2	Total	Heat Pump Portion	Supplementary Resistors Portion 3	Annual Diversified Load Factor, Per Cent
Heat Pump Systems											
A. With supplementary resistors; night set-back of indoor thermostat	2	18.1	6.8	11.3	12.6	6.8	4.0	18.1	15.8	2.3	16.4
	3	15.8	4.5	11.3	10.4	4.6	2.7	12.8	10.5	2.3	14.0
	4	14.7	3.4	11.3	9.3	3.4	2.0	10.2	7.9	2.3	12.5
	5	14.0	2.7	11.3	8.7	2.7	1.6	8.6	6.3	2.3	11.3
B. With supplementary resistors; fixed setting of indoor thermostat	2	12.0	6.0	6.0	8.2	6.8	4.0	17.7	16.3	1.4	24.6
	3	10.0	4.0	6.0	6.0	4.6	2.7	12.2	10.8	1.4	23.2
	4	9.0	3.0	6.0	4.9	3.4	2.0	9.5	8.1	1.4	22.1
	5	8.4	2.4	6.0	4.2	2.7	1.6	7.9	6.5	1.4	21.5
C. With supplementary heat storage	2	8.0	8.0		6.8	6.8	4.0	18.7	15.3	3.4	31.4
	3	5.3	5.3		4.6	4.6	2.7	13.6	10.2	3.4	33.8
	4	4.0	4.0		3.4	3.4	2.0	11.0	7.7	3.4	37.3
	5	3.2	3.2		2.7	2.7	1.6	9.5	6.1	3.4	40.2
Resistance Heating System, Including Electric Cooling (4)											
D. With night set-back of indoor thermostat		20.9			18.0	12.2	2.9	30.6			19.4
E. With fixed setting of indoor thermostat		16.1			13.1	22.2	2.9	30.6			26.7
F. With supplementary heat storage		9.8			9.8	9.8	2.9	30.6			35.6

* Eastern standard time. † Eastern daylight time.

1. Maximum diversified demands and diversified demands for winter conditions are for an average day of coldest spell.
2. Diversified demands for summer conditions are for an average day of warmest spell.
3. Includes energy use for defrosting.
4. Separate cooling system reflects a 20% better efficiency than present-day cooling equipment.

mentary resistors fall far short of meeting the reference band in the present-day range of annual COP. But the type with supplementary resistors and fixed setting of indoor thermostat, reaches it in the very high region of annual COP.

3. All three types of resistance systems fall far short of meeting the reference band.

4. In the region of annual COP of 3, the difference in the cost to serve the heat pump systems with supplementary heat storage and to serve those with supplementary resistors is in the order of 70×10^3 cost units with respect to type *B*, and 350×10^3 cost units with respect to type *A*. These differences represent 17% and 52% respectively of the annual cost to serve the loads of these two types of heat pump systems.

Cost Rate per Kw-Hr

Fig. 2(B) depicts the relationships of the cost-to-serve rates per kw-hr, with changing annual COP. The following important features will be noted.

1. Only the heat pump with supplementary heat storage provides an ever-declining cost-to-serve rate per kw-hr. Heat pumps with supplementary resistors reflect the reverse phenomenon; their cost-to-serve rates per kw-hr rise appreciably with an improvement in the annual COP.

2. The cost-to-serve rates per kw-hr of resistance systems are fixed. They lie in the range between $2^1/_2$ and 4 cost units. For the system with supplementary heat storage the cost rate approaches that of the heat pump with supplementary heat storage at an annual COP of 3.

These features in the cost-to-serve rates per kw-hr should be of extreme significance to utility managements in establishing pricing policies for the supply of these loads, not only in respect to the level but, more importantly, in respect to the proper form of rates. A simple energy rate per kw-hr, for example, now generally in use on electric utility systems in the United States for service supplied to residential customers, cannot adequately meet the conditions depicted in Fig. 2(B) not only as to various types of heating systems, but also as to probable future improvements in the annual COP. To illustrate, if the price per kw-hr of such a rate is set to satisfy, say, the type *A* system at a given annual COP, it would be far too high for the types *B* and *C* at the same annual COP, and too high for any of the resistance-type systems. On the other hand, if the price per kw-hr of such a rate is set to satisfy the requirements of type *C* system at a given annual COP, it would be too low to satisfy the requirements of types *A* and *B* at that COP, and too low to satisfy the requirements of resistance types *D* and *E* systems and even of type *F*, if the rate is set for a COP region beyond $3^1/_2$. Even more significant, if a simple energy type rate were set to satisfy, say, the requirements of either type *A* or *B*, under the now existing low levels of heat pump COP, it would become inadequate in the future for an improvement in their COP.

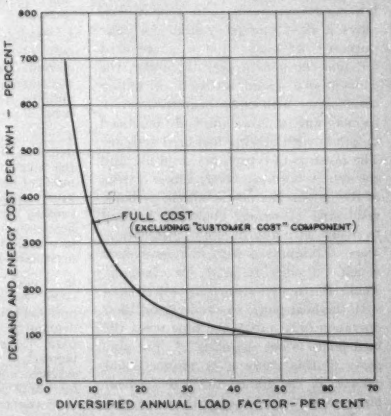


Fig. 1. Relation of demand and energy cost per kw-hr and annual diversified load factor, 100% full cost per kw-hr, at 45% annual load factor

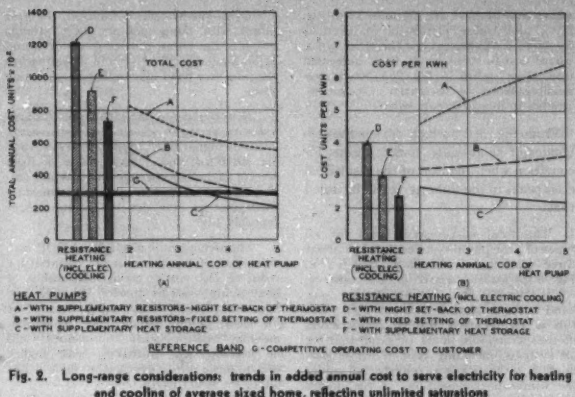

TOTAL ANNUAL COST UNITS x 10³
TOTAL COST
COST PER KWH
COST UNITS PER KWH
RESISTANCE HEATING (INCL ELEC COOLING)
HEATING ANNUAL COP OF HEAT PUMP
(A)
(B)
HEAT PUMPS
A - WITH SUPPLEMENTARY RESISTORS - NIGHT SET-BACK OF THERMOSTAT
B - WITH SUPPLEMENTARY RESISTORS - FIXED SETTING OF THERMOSTAT
C - WITH SUPPLEMENTARY HEAT STORAGE
RESISTANCE HEATING (INCL ELECTRIC COOLING)
D - WITH NIGHT SET-BACK OF THERMOSTAT
E - WITH FIXED SETTING OF THERMOSTAT
F - WITH SUPPLEMENTARY HEAT STORAGE
REFERENCE BAND G - COMPETITIVE OPERATING COST TO CUSTOMER

Fig. 2. Long-range considerations: trends in added annual cost to serve electricity for heating and cooling of average sized home, reflecting unlimited saturations

This is the only residential utilization device known to the author which would require an increase in the price per kw-hr with an improvement in its efficiency of performance.

Thus, it behooves those responsible for pricing policies of utility services to bear in mind these important and fundamental economic features of electric heating loads, and especially those of the heat pumps with supplementary resistors. As has been pointed out in reference 4, with the wide variation possible in the relative cost to serve various types of heat pump, resistance heating, cooling, and other loads in the residential field, together with the wide variety of operating conditions that may be encountered in the characteristics of these loads, adequate pricing for the all-electric home, equitable to various types of residential customers and to utilities, can be accomplished only through some form of "demand-energy" type of rates. It could be met reasonably well also through a form which would employ some kind of dual-dial metering of energy, or its equivalent, by applying a lower price for the kw-hr registered during certain specified hours of the day, as contrasted with a much higher price for either all, or the excess over a given base, of kw-hr registered during the remaining or peaking, hours of the day, as has been suggested in reference 10.

Short-Range Economic Considerations

Departing now from the fundamental, or long-range, considerations to those which are temporarily affected by the load characteristics of the residential class of service of which the loads of the heating systems are a part, the economic aspects of these loads, from short-range considerations, are found to differ from those which have been described for the long-range. This is due to two factors: first, as has been pointed out in reference 4, the total load of the residential class of customers, as it exists at any given time, contains peaks and valleys in relation to its morning and evening loads in the winter, and in relation to its seasonal peak patterns between summer and winter; and second, the load of heating systems with their complementary cooling possesses its own peaks and valleys and its own seasonal patterns. These characteristics of the heating and cooling loads thus can be fitted into the load patterns of the residential class. However, these favorable conditions exist only for relatively small saturation levels of these loads.[4] Thus, from such short-range considerations, the economic aspects fall into two categories: 1. those applying to conditions when the total load of the residential class peaks in the winter, and 2. those applying to conditions when that load peaks in the summer.

Since, for winter conditions, the present-day residential class peak occurs in the evening, while the maximum load of heating occurs in the morning,[11] the resulting load diversity permits the acquisition of a limited amount of the heating load at lower demand cost responsibility than applies to its morning peak. For summer peak conditions, the summer demand of a combination heating and cooling system on the average day of the warmest spell is lower than on the average day of the coldest spell in the winter.[11] The resulting load diversity here also permits the acquisition of a limited amount of this load at a lower demand cost responsibility than applies to its winter peak.

Fig. 3(A) and (B) depict the cost-to-serve relationships, under the aforementioned lower demand cost responsibilities, and show the limits of saturation which would be permissible under these short-range considerations. The cost curves were computed with the use of the load characteristics data shown in Table I and diversified demands applicable to the winter evening or to the summer evening conditions, as the case might be. In making these computations use was made

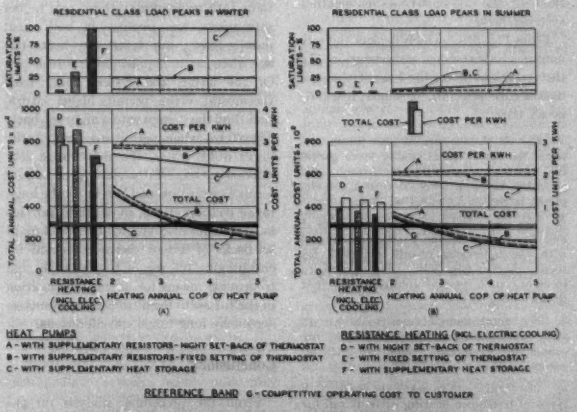

RESIDENTIAL CLASS LOAD PEAKS IN WINTER
RESIDENTIAL CLASS LOAD PEAKS IN SUMMER
SATURATION LIMITS - %
TOTAL ANNUAL COST UNITS x 10³
COST UNITS PER KWH
TOTAL COST
COST PER KWH
RESISTANCE HEATING (INCL ELEC COOLING)
HEATING ANNUAL COP OF HEAT PUMP
(A)
(B)
HEAT PUMPS
A - WITH SUPPLEMENTARY RESISTORS - NIGHT SET-BACK OF THERMOSTAT
B - WITH SUPPLEMENTARY RESISTORS - FIXED SETTING OF THERMOSTAT
C - WITH SUPPLEMENTARY HEAT STORAGE
RESISTANCE HEATING (INCL ELECTRIC COOLING)
D - WITH NIGHT SET-BACK OF THERMOSTAT
E - WITH FIXED SETTING OF THERMOSTAT
F - WITH SUPPLEMENTARY HEAT STORAGE
REFERENCE BAND G - COMPETITIVE OPERATING COST TO CUSTOMER

Fig. 3. Short-range considerations: trends in added annual cost to serve electricity for heating and cooling of average sized home and permissible limits of saturation

of the cost-to-serve relationship shown in Fig. 1 under the same assumptions as were used in preparing Figs. 2(A) and (B), except for the application of an additional cost factor to reflect adequately the customer demand component of the basic cost equation described in references 5 and 7. Under the lower diversified demands here employed, the cost relationship of Fig. 1 does not reflect fully this component of costs, since it is based on the maximum of such demands.

The curves for permissible saturations were developed from data of the residential class of service of the metropolitan electricity supply system previously described, for the foreseeable magnitude of margins available between the winter period evening peak and the morning load; and between the estimated summer and winter peaks a decade after the class peak were to shift to the summer period. Although different numerical results would be obtained for a set of different assumptions, those used herein are believed to be within practical possibilities. It must be noted, however, that the magnitudes of these margins do not remain fixed over a period of time because of the continuous ebb and flow in the use of electricity by customers and the constantly changing patterns of that use, as set forth in reference 12. Such margins may either increase or decrease as time goes on. Thus any computations made on the basis of foreseeable conditions possess elements of chance as to the probable future behavior of the load patterns of a whole class of customers, and therefore should be viewed in a broad sense as indicative of the relative significance in the permissible saturations of the various types of heating systems shown.

From Fig. 3(A), which depicts the economics of short-range considerations for the conditions when the load of the residential class peaks in the winter, it will be noted that:

1. For all three types of heat pump systems, the total annual cost-to-serve curves fall within a relatively narrow range, and approach the cost level of type *C*, as depicted in Fig. 2(A).

2. The cost per kw-hr curves of types *A* and *B* systems become fixed as contrasted with the increasing trends depicted for these types in Fig. 2(B). The downward trend of type *C* remains unaltered from that shown in Fig. 2(B).

3. The total annual costs to serve all three types of heat pump systems reach the competitive reference band around an annual COP of 3.

4. The total costs to serve the load of all types of resistance heating systems fall far short of the competitive reference band.

5. Except for types *C* and *F* systems, whose cost to serve are the same as in Fig. 2(A), and whose permissible saturations therefore are unlimited, for the other types, whose cost-to-serve levels have decreased from those shown in Fig. 2(A), their permissible limits of saturation become restricted to relatively low values.

From Fig. 3(B), which depicts the economics of short-range considerations for condition when the load of the residential class peaks in the summer, it will be noted that:

1. The total cost-to-serve level for all types of heat pump systems drops to about three quarters of the level depicted in Fig. 3(A) for the winter peak conditions. All of them reach the competitive reference band around annual COP of $2^1/_2$. But this is accompanied also by a restriction in the permissible limits of saturation for all systems, including type *C* whose permissible saturation was unrestricted under the winter peak conditions at the higher costs.

2. Similarly, the total cost-to-serve level for all types of resistance heating systems drops to one-half of the level depicted in Fig. 3(A), for the winter peak conditions. But even at this low level, neither of them reach the competitive reference band; and their permissible saturation limits decrease to insignificance.

Thus, from short-range considerations also, it behooves those responsible for pricing policies of utility services to realize that the lower costs to serve are feasible for electric heating systems only during the period while their load is filling valleys which exist in the hourly and seasonal patterns of the residential class load. Once these valleys are filled, additional load must bear the full and higher demand cost responsibilities. Higher saturations are permissible with heat pumps than with resistance systems, and higher saturations are permissible with heat pumps having supplementary heat storage than with supplementary resistors. Of the latter, the system using fixed thermostat setting permits higher saturation than the system with a night set-back. Therefore, while permitting temporarily lower utility rates, the short-range considerations require the same type of rate forms as were outlined herein for the long-range considerations, in order to provide incentives for discouraging the use of night set-backs of indoor thermostats and for encouraging the use of heat storage, and to provide means for an easier transition in the future to the ultimate conditions reflected by long-range considerations.

Conclusions

From the preceding analysis on the basis of long-range and short-range considerations it can be concluded that:

1. Of all the systems considered, the heat pump with supplementary heat storage provides the lowest cost-to-serve level, and at an annual COP of around 3 approaches the reference band of competitive operating costs.

2. Heat pump systems with supplementary resistors produce acceptable cost-to-serve levels at higher COP than 3, provided they are sized for, and operate under, a fixed setting of indoor thermostats. When employing night set-back of such thermostats their cost-to-serve levels become unacceptably high.

3. Improvements in the annual COP into the region of 3 and higher are essential for reducing the cost-to-serve relationship of heat pump systems toward the reference band of competitive operating costs. But for heat pumps which employ supplementary resistors such improvement will be accompanied ultimately with higher cost-to-serve rates per kw-hr than exists under the lower COPs.

4. The cost-to-serve levels of any of the resistance-type systems described fall far short of meeting the reference band of competitive operating costs. Their permissible saturation limits become insignificantly low under the short-range considerations as a possible seasonal valley-filling load.

Thus, if an all-electric home is to become a mass-market reality, there must be further intensification in research and development of the heat pump for substantial improvement in its annual COP, and the adoption of some form of supplementary heat storage, which is dependent upon a successful development of heat storage means. To provide incentives toward encouraging these requirements, and discouraging the use of supplementary resistors in conjunction with the heat pump, especially under night set-backs of indoor thermostat settings, and to facilitate meeting the economic necessities of long-range considerations, electric service to all types of house heating systems, with their complementary cooling, should be priced under some form of "demand-energy" rates, or its equivalent, because this type of rate form can be made to meet, over a wide range of load factors, the fundamental hyperbolic nature of the cost-to-serve versus load factor relationship such as is depicted in Fig. 1.

References

1. The Heat Pump—Its Significance as a Potential Residential Electric Load. **Constantine Barry.** *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 1–8.

2. The Heat Pump Load—Good or Bad? **Constantine Barry.** *Refrigerating Engineering*, American Society of Refrigerating Engineers, New York, N. Y., May 1950, pp. 451–55.

3. Residential Heat Pump Experiments in Philadelphia—Suggested Possibilities for

PRACTICAL APPLICATIONS, **Constantine Bary**, *AIEE Transactions*, vol. 71, pt. II, 1952 (Jan. 1953 section), pp. 361–65.

4. LOAD AND ECONOMIC ASPECTS OF THE RESIDENTIAL HEAT PUMP ON ELECTRIC UTILITY SYSTEMS, **Constantine W. Bary**, *Ibid.*, vol. 76, pt. II, May 1957, pp. 49–54.

5. ECONOMIC SIGNIFICANCE OF LOAD CHARACTERISTICS AS APPLIED TO MODERN ELECTRIC SERVICE, **Constantine Bary**, *Minutes*, 53rd Annual Meeting, Association of Edison Illuminating Companies, New York, N. Y., 1937 (printed but not published).

6. LOAD-STRUCTURE OF A MODERN ELECTRIC UTILITY SYSTEM, **Constantine W. Bary**, *Proceedings*, American Power Conference, Chicago, Ill., vol. XVI, also *Electric Light and Power*, Chicago, Ill., May, 1954.

7. LOAD CHARACTERISTICS OF "SMALL LIGHT AND POWER" CUSTOMERS—THEIR ECONOMIC SIGNIFICANCE, **Constantine W. Bary**, *Electric Light and Power*, Chicago, Ill., Oct. 15, 1955.

8. REPORT ON HEAT STORAGE, **Joint AEIC-EEI Heat Pump Committee**, *Bulletin*, Edison Electric Institute, New York, N. Y., Aug. 1958, p. 261.

9. HEAT PUMP APPARATUS, **Constantine W. Bary**, *U. S. Patent no. 2,723,083*, Nov. 8, 1955.

10. *Discussion* by **Constantine Bary**, *Minutes*, 65th Annual Meeting, Association of Edison Illuminating Companies, 1949, p. 80.

11. COOLING LOAD NEEDS COMPLEMENT, **Constantine W. Bary**, *Electrical World*, New York, N. Y., May 9, 1955.

12. COINCIDENCE-FACTOR RELATIONSHIPS OF ELECTRIC-SERVICE-LOAD CHARACTERISTICS, **Constantine Bary**, *AIEE Transactions*, vol. 64, 1945, pp. 623–28.

Discussion

J. Grant De Remer (Consulting Engineer, Darien, Conn.): This paper is vitally important in showing the marked distinction between short-range and long-range considerations. Throughout the history of electrical utility operation the industry has suffered the penalty involved in supplying peak power at midwinter. Until recently, the peak usually occurred on the Saturday night before Christmas.

I have examined load curves of large utilities covering the past 40 years, many of which show that because of this fact, from one quarter to one third of the total plant and distribution system investment has been involved in the production of only about 1% of the annual output. The obvious high fixed charges resulting from this condition have always been a challenge to the electrical engineering profession.

Now, however, with the coming of summer air conditioning and large irrigation pumping, particularly in the western states, some utilities are experiencing maximum annual peaks in the summer. It is these conditions which Mr. Bary so clearly sets forth in this paper.

Until the time in the not distant future when the new peak from summer air conditioning exceeds the winter peak in a system, the cost and rate factors involved take on the characteristics of his short-range analysis. As the long-range period approaches, Mr. Bary brings out clearly the importance of storage, especially heat storage, in connection with the use of the heat pump. The possibilities in this direction are very great, especially when they are considered in conjunction with solar heat sources.

The vital feature needed here is the discovery of a practical, low-cost heat storage medium, preferably one that is applicable to short-hour, high-temperature, daily off-peak storage, sufficient to carry a residence over the off-peak hours, as is done with electric water heaters, plus a longer-hour, low-temperature heat storage in the 40- to 60-degree Fahrenheit range.

The magnitude of this latter storage will be determined by the extreme weather conditions in the latitudes concerned. Major attention, therefore, should be given by research organizations into the possibilities of these types of heat storage. The thinking in this direction is comprehended by Mr. Bary's reference to "new approaches," as a radical breakthrough seems to be needed at this point.

L. P. Herrington (John B. Pierce Foundation, Inc., New Haven, Conn.): Mr. Bary's analyses serve to reinforce certain earlier conclusions on short-range effects and to extend information on long-range effects. Of the many features that may be deduced from his data, the following items probably need wider appreciation among groups concerned with heat pump design and manufacture rather than utility operation:

1. The summer margin which may be filled with a seasonal load such as cooling is an area which, when filled, changes the economic status of the cooling load from one of partial cost to a full-cost basis. In view of the hyperbolic relation between diversified annual load factor and demand-and-energy cost per kw-hr, the full-cost basis above the marginal area exceeds the partial cost per kw-hr by a factor as large as seven at low diversified annual loads. This is in accord with the established principle that cost-to-serve on large loads such as house heating must be considered with respect to maximum diversified demand as well as with respect to short-term effects on the load characteristics of a specified class of service.

2. Since the competitive aspects of sales of heating as contrasted with cooling service are markedly different for electric systems, it is important to know the effect of annual heat pump system COP on this factor. The data indicate that although improvement in COP has a favorable effect on cost-to-serve, it is only the heat pump system with supplementary heat storage that approaches competitive status with an annual COP of 3.

3. A further fact which characterizes the heat pump with supplementary heat storage as contrasted with supplemental resistance heat is that systems with heat storage provide a steady decrease in cost-to-serve in contrast to the increasing costs of resistance heat supplement.

4. Improvement in cost-to-serve factors with heat pump systems are noted for the condition of fixed as opposed to night set-back thermostat operation, design factors being assumed adequate for this operation. Consideration of this understandable conclusion ought to be influenced by two factors: 1. the common impression of the fuel savings achieved by night thermostat set-back is probably greater than is actually the case, and 2. the night reduction in thermostat level achieves cooler sleeping quarters, without drastic ventilation, a performance feature valued by many.

5. Mr. Bary's highly informed study is probably widely applicable, his northeastern metropolitan system being reasonably typical of future demand. Generalization of his equations and data on costs, general system operating characteristics, and specific equipment characteristics, which are currently to some degree dimensionally distinct, might present a formidable problem but would be in line with the recent extension in the feasibility of practical computation, and would provide a valuable guide for policy in electric systems differently situated.

D. W. Locklin and **J. A. Eibling** (Battelle Memorial Institute, Columbus, Ohio): Electric-heating specialists generally recognize that night set-back of the thermostat is undesirable with electric space-heating systems. However, it is doubtful that there is always appreciation of the magnitude of the penalty inflicted by night set-back if extended to high saturation. The paper forcibly shows the need to discourage night set-back of the thermostat.

Another highly revealing point demonstrated by the analysis is that the annual diversified load factor decreases appreciably with improvement in COP when supplementary resistance heating without heat storage is used. Thus, with the present practice of using supplementary resistance heating, further improvements in heat-pump performance made by manufacturers can actually result in higher average power cost per kw-hr if such systems are installed in substantial saturation. Conversely, it is particularly interesting to learn of the great extent to which supplementary heat storage would reduce the demand factor of the electric rate structure. This attests to the desirability of having as soon as possible a reliable, low-cost storage system for residential application.

Recently, an exploratory study of heat storage as applied to residential heat pumps was conducted at Batelle for the Joint Association of Edison Illuminating Companies–Edison Electric Institute Heat Pump Committee. This study revealed that the development of compact and reliable heat storage is feasible with the use of high-temperature storage media charged by resistance heaters. In application, the high-temperature storage would be charged during periods controlled by a time clock or demand switch, in such a way as to provide desirable load characteristics. The utilization of the heat of fusion of certain salt hydrates has received widespread attention for lower temperature storage, but their usage is faced with difficulties, mainly associated with erratic crystallization behavior, yet to be overcome.

From Mr. Bary's analysis, the development of practical and economical heat storage systems is seen to be an extremely imporant factor in the broadening of the electric-space-heating market. As a corol-

lary to this, an incentive such as a lower power rate must be offered to purchasers of residential heat pumps to encourage their investing in supplementary heat storage in preference to supplementary resistance heating. Careful analyses such as this are an important adjunct toward defining the direction of research and development.

Constantine W. Bary: I am grateful to the discussers for their emphasis on one or more of the aspects contained in the paper for the achievement of the all-electric home as a mass-market reality. Their thoughts should aid in bringing about better realization of the concerted actions needed in the fields of the following:

1. Intensive research and development for the use of economically acceptable supplementary electrical heat storage means, to minimize the cost to serve the electrical load of an all-electric home.

2. Continued research and development for substantial improvements in the COP of the electrical heat pump, to bring about costs of heating with electricity closer to the competitive reference band.

3. Establishment of demand-energy rates for such applications to insure the relative adequacy of electricity prices over a wide range of conditions, and to provide self-acting economic incentives for encouraging the use of supplementary heat storage and for discouraging the use of night set-back of indoor thermostats.

All three of these actions are required at the utilization level of electric systems' operations. Mr. De Remer mentions the challenges presented by short-duration system peaks. I foresee that, in the long-range aspects, and for high degree of saturation of all-electric homes, such conditions will continue to present challenges. Therefore, actions will be required also at the supply level of electric systems' operations. As pointed out in reference 4, this could take the form of the relatively new idea to inject periodically into the growing capacity requirements of a supply system a certain amount of (now considered unconventional) low first-cost peak-shaving generating equipment such as the package type, unattended diesel engine or gas-turbine plants for decentralized operation throughout the distribution system at times of peaks and for reserves.

Signal Stabilization of a Control System

R. OLDENBURGER
MEMBER AIEE

C. C. LIU
NONMEMBER AIEE

FEEDBACK loops, in states of self-oscillation, can often be stabilized by the introduction of an appropriate signal. In this paper a theory is developed to explain this phenomenon. The theory is applied to a system which is linear except for a limiter. Experimental results for the system showing the removal of the hunt by a stabilizing signal are given in a previous paper.[1]

For nonlinear elements with the output a function of the input a gain concept is defined so that these nonlinear elements may be replaced by simple amplifiers with this type of gain. Describing functions are used to relate sinusoidal stabilizing signals to the corresponding inputs and outputs of the limiter. This relation is described by the points of intersection of two families of curves with the frequency and amplitude of the stabilizing signal as parameters. From these curves the equivalent gain of the limiter is readily obtained, given the constants of the system and the stabilizing signal. The characteristic roots of the system are then chosen to give good transient response. Because of the general availability of 60- or 400-cycle-per-second electric power, signal stabilization can often be easily introduced in industrial applications.

Stabilization

Automatic control systems are often stabilized by the use of a properly chosen linear band of operation about equilibrium. Thus control systems that are otherwise nonlinear have been stabilized with the aid of a linear band in the sense that when the errors in the controlled variable and its derivatives are small enough, the system behaves like a linear system whose characteristic roots are satisfactory. Stability here is achieved by nonlinear means.

The phenomenon of this paper was discovered experimentally by one of the authors during frequency response runs on a hydroelectric system simulated on an analog computer.[1] The stabilization dealt with here depends on the introduction of an appropriate signal which increases or decreases the effective open-loop gain of the system in the neighborhood of equilibrium, without materially affecting the gain for large deviations from equilibrium. It is assumed, as is usually the case, that the self-oscillations of the system in the absence of the stabilizing signal are bounded. It is also assumed that the hunt is not completely wild, but that a moderate reduction in open-loop gain will remove the hunt. The stabilizing signal may then leave the response to large disturbances relatively unaffected while at the same time removing the hunt. Signal stabilization is analogous to, but not identical with, the insertion of an additional amplifier or attenuator into the given loop. Because of the general availability of 60-cycle-per-second electric power, the introduction of signal stabilization is often a simple matter. This paper is devoted to a method of analysis for computing stabilizing signals.

MacColl showed that the introduction of a sinusoidal signal at the input to an on-off servomechanism yields a system that behaves like a linear one for small inputs superimposed on this signal.[2] This differs from the illustrative example of the present paper in that the system of this paper is already linear in the neighborhood of equilibrium, whereas the system of MacColl is not.

Minorsky treated the use of a signal to excite or quench the hunt (self-oscillation) of a physical system described by a second-order differential equation in the variable $x(t)$, for time t, where the coefficient of the "damping" term in dx/dt is an even polynomial in x.[3] The stabilization of Minorsky depends on properties of asymptotic solutions of the differential equation. The stabilization technique of this paper does not depend on these properties.

System With Bounding Element

In every problem involving a hydraulic servomotor the speed of the servomotor piston is bounded. In Fig. 1 the system to be controlled is shown in one block.

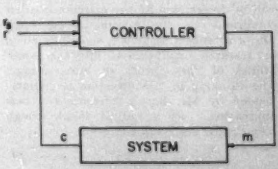


Fig. 1. System to be controlled

Paper 59-219, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 20, 1958; made available for printing December 10, 1958.

R. OLDENBURGER is with Purdue University, Lafayette, Ind., and C. C. LIU is with the Taylor Instrument Company, Rochester, N. Y.

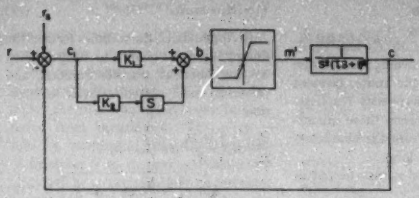


Fig. 2 (left). Nonlinear control system

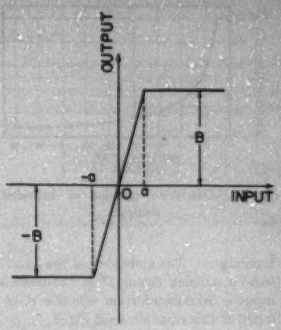


Fig. 3 (right). Characteristic of a bounded element

The variable m is the controlling variable, such as the throttle on a gasoline engine. If the controller is a governor, m can be taken to be servomotor piston position. The output of the controlled system is denoted by c. For a governor-engine system, c is engine rpm (revolutions per minute). The input to the controller is $r-c+r_s$, where r is the reference value of the controlled variable (speed setting in the case of a governor-engine system), and r_s is the value of the extra input signal used to stabilize the system. Stabilization with r_s will be called "signal stabilization." This signal need not be introduced at the same point as the reference r, but for simplicity this will be assumed here.

The theory of signal stabilization will be developed in this paper in connection with a controlled system possessing the transfer function

$$\frac{1}{s^2(\tau s+1)^2}$$

preceded by a limiter as shown in Fig. 2. The limiter follows a proportional plus derivative controller with gains K_1 and K_2 in the proportional and derivative channels respectively.

Equivalent Gain of Nonlinear Element

Let the input and output to a nonlinear element be denoted by x_1 and x_2 respectively, and assume that

$$x_2=f(x_1) \tag{1}$$

for a function $f(x_1)$, whence the nonlinear element is time independent. Let S^* denote a periodic stabilizing signal with average value equal to zero. It is assumed that when S^* is the entire input to the nonlinear element, in which event $x_1=S^*$, the average of the output of the element is zero. A constant second input component K_b is added to the stabilizing signal S^* to yield an input $x_1=S^*+K_b$. Let the average of the corresponding output of the nonlinear element be denoted by A_v. The equivalent gain g_b of the nonlinear element is defined by the following:

$$g_b=\lim_{K_b\to 0}\frac{A_v}{K_b} \tag{2}$$

The equivalent gain is thus the limit of the average value of the output divided by the average value of the input as the average value of the input goes to zero.

For sinusoidal stabilizing signals S^* and odd functions $f(x)$ [$f(x_1)$ is odd when $f(-x_1)=-f(x_1)$] the limit process of equation 2 can be carried out. Such functions describe many of the commonly occurring nonlinearities such as limiting and dead band with the elements otherwise proportional.

The input-output characteristic of a limiter is shown in Fig. 3. The relation between the input x_1 and the output x_2 of the limiter is given for constants a and B by

$$\begin{aligned} x_2&=\frac{B}{a}x_1, && -a<x_1<a\\ x_2&=B, && x_1>a\\ x_2&=-B, && x_1<-a \end{aligned} \tag{3}$$

Let the input to the limiter be the sinusoidal wave given by

$$x_1=K_b+B_1\sin\omega t \tag{4}$$

for the positive K_b. Let

$$B_1>a+K_b \tag{5}$$

so that the input x_1 reaches both corners of the characteristic curve of the limiter. As proved in Appendix I, the equivalent gain g_b for the limiter is given by

$$g_b=\frac{B}{a}\frac{1}{\pi}\left[2\left(\frac{a}{B_1}\right)+\frac{1}{3}\left(\frac{a}{B_1}\right)^3\right] \tag{6}$$

A curve for equation 6 is plotted in Fig. 4 for the case where $a=0.2$ and $B=10$.

If a sinusoidal stabilizing signal r_s is introduced into a feedback loop with a nonlinear element, the resulting input to this element will not in general be sinusoidal. In this event, as in describing function theory, the fundamental component of the periodic wave is employed.

System Stability

The limiter in the feedback loop of Fig. 2 is replaced by a simple amplifier with the gain g_b of equation 6. The characteristic equation of the system is given by

$$\tau^2s^4+2\tau s^3+s^2+K_2g_bs+K_1g_b=0 \tag{7}$$

By Routh's stability criterion this equation is stable if and only if

$$g_b<\left(\frac{2}{K_2}-\frac{4K_1}{K_2^2}\right) \tag{8}$$

In view of the inequality of equation 8, if the equivalent gain is high enough, the loop will experience sustained oscillations. The radian frequency ω_n for sustained oscillations is given by

$$\omega_n=\frac{1}{\tau}\sqrt{1-2\tau\frac{K_1}{K_2}} \tag{9}$$

From K_1, K_2, and τ the equation

$$g_b=\left(\frac{2}{\tau K_2}-\frac{4K_1}{K_2^2}\right)$$

yields the critical value of g_b for such

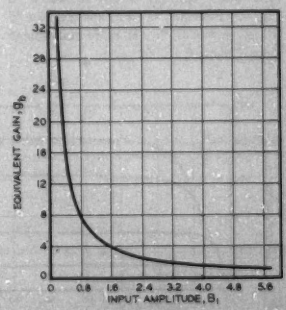


Fig. 4. Equivalent gain of bounded element

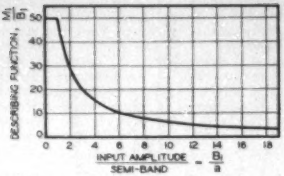


Fig. 5. Describing function of bounded element

oscillations. The amplitude of the sinusoidal stabilizing signal S^* at the limiter input is determined from equation 6, or a plot of this equation as in Fig. 4.

Describing Function

The stabilizing signal r_s that will yield the desired stabilizing signal at the input to the limiter of Fig. 2 will now be established with the aid of describing functions.

Let x_1, for

$$x_1 = B_1 \sin \omega t \tag{10}$$

be the input to the limiter, and

$$M_1 \sin \omega t \tag{11}$$

the fundamental component of the output x_2 of the limiter. The describing function[4] of the limiter is M_1/B_1, where

$$\frac{M_1}{B_1} = \frac{B}{a}\frac{2}{\pi}\left[\sin^{-1}\left(\frac{a}{B_1}\right) + \left(\frac{a}{B_1}\right) \times \sqrt{1-\left(\frac{a}{B_1}\right)^2}\right] \tag{12}$$

A plot of formula 12 is given in Fig. 5.

Let the stabilizing signal r_s be given by

$$r_s = R_s \sin \omega t \tag{13}$$

for the amplitude R_s of r_s. A modified plot of the describing function for the case where $a=0.2$, $B=10$ is shown in Fig. 6. For what is to follow the plot of M_1/R_s versus B_1/R_s is more convenient than a plot of M_1/B_1 versus B_1/a.

Frequency Response

A relation between M_1/R_s and B_1/R_s was obtained in the preceding section from the characteristics of the limiter element only with the aid of frequency response. A second relation between these quantities can be obtained from the characteristics of the rest of the loop.

The stabilizing signal r_s may be put in the complex form

$$R_s e^{j\omega t}$$

for $j=\sqrt{-1}$. Suppose that the input r to the loop is zero, and that the loop is stable when r_s is introduced. The complex forms of the fundamental components of the input and output of the limiter are taken to be

$$B_1 e^{j(\omega t+\theta)}, \quad M_1 e^{j(\omega t+\theta)}$$

respectively, for a phase angle θ. The complex form C of the fundamental component of the output c of the loop is given by

$$C = -M_1 H_1 e^{j(\omega t+\theta-2\alpha)} \tag{14}$$

where

$$\alpha = \tan^{-1} \omega\tau$$

$$H_1 = \frac{1}{\omega^2(\omega^2\tau^2+1)}$$

The complex form C_1 of the fundamental component of the controller input c_1 is given by

$$C_1 = [R_s + M_1 H_1 e^{j(\theta-2\alpha)}]e^{j\omega t} \tag{15}$$

In Appendix II, the complex form of the limiter input is equated to C_1 multiplied by the controller transfer function, and the result manipulated to yield the equation

$$\frac{B_1}{R_s} = H_2\left\{\frac{M_1 H_1}{R_s}\cos(2\alpha-\beta) \pm \sqrt{1+\left[\frac{M_1 H_1}{R_s}\sin(2\alpha-\beta)\right]^2}\right\} \tag{16}$$

For given system constants, the parameters H_1, H_2, α, and β are fixed. With R_s also given equation 16 yields a relation between B_1 and M_1.

The two relations between B_1 and M_1 determine these quantities when the stabilizing signal and the loop constants are given. The dependence of the stabilizing signal S^* at the limiter input on the system stabilizing signal r_s is now determined.

Relation Between Stabilizing Signals

In Fig. 7 a family of curves relating M_1/R_s and B_1/R_s is given with R_s as the parameter. These curves depict equation 12. On this same figure is plotted a family of curves satisfying equation 16 with ω as the parameter.

A point of intersection of a curve of one family with a curve of the other determines the amplitude of B_1, as well as M_1.

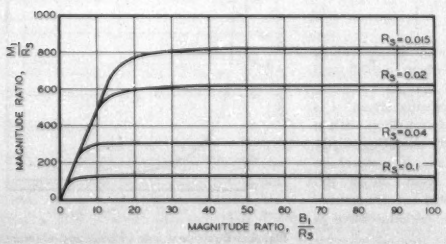


Fig. 6 (below). Modified describing function of bounded element

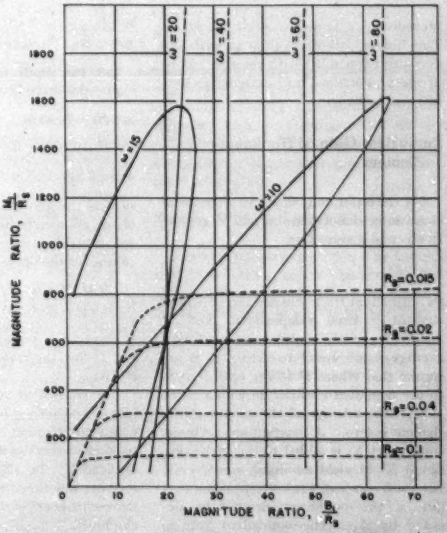


Fig. 7 (right). Diagram for obtaining B_1, for a given R_s and ω

For $\omega > 15$ there is only one such intersection point. In this case the amplitude of B_1, and hence M_1 is uniquely determined from the figure. With B_1 determined the equivalent gain g_b can be computed from equation 6. This may in turn be substituted into the inequality of equation 8 to determine system stability. The values of g_b must be restricted to an appropriate range so that the transient response will be satisfactory.

If the dominant roots of the system characteristic equation 7 are of the form $-P \pm jQ$, $0 \leq P \leq Q \leq 2P$, the transient response is good, provided that P is large enough. Given P, and if Q dominates $2P$, the system response is too oscillatory, whereas if P dominates Q, the deviations in the output are excessive.

Numerical Case

Experimental results showing the removal of a hunt by the introduction of a stabilizing signal are given in a previous paper[1] for the system of Fig. 2 with

$K_1 = 1$
$K_2 = 0.8$
$B = 10$
$a = 0.2$
$\tau = 0.1$

The characteristic equation for this system is given by

$$x^4 + 20x^3 + 100x^2 + 80g_b x + 100g_b = 0 \tag{17}$$

For system stability it is necessary that

$$g_b < 18.75 \tag{18}$$

The gain 50 of the limiter for zero input is higher than this limit. Therefore in physical tests on the system, the loop went into self-oscillation. Equation 9 gives the frequency of the hunt, which is

$\omega_n = 8.66$ radians per second

From the characteristic equation 17, the inequaly 16, and certain other considerations, it can be proved, as shown in Appendix III, that the values of equivalent gain which yield good transient response are in the range

$$1.6 < g_b < 4 \tag{19}$$

This explains the experimental run of Figs. 5 and 6 in the previous paper.[1] The range of the values of the amplitude B_1 of the fundamental component of the input to the limiter corresponding to the inequalities 19 is

$$1.56 < B_1 < 3.9 \tag{20}$$

Appendix I. Equivalent Gain for the Limiter

Let the input x_1 to the limiter of Fig. 3 be given by

$$x_1 = K_b + B_1 \sin \omega t \tag{4}$$

where

$$B_1 > a + K_b \tag{5}$$

The average value A_y of the output x_2 corresponding to this input x_1 is then

$$A_y = \frac{1}{2\pi}\left[\int_0^{\lambda_1} \frac{B}{a}(K_b + B_1 \sin \omega t)d\omega t + \int_{\lambda_1}^{\pi-\lambda_1} B d\omega t + \int_{\pi-\lambda_1}^{\pi+\lambda_2} \frac{B}{a}(K_b + B_1 \times \sin \omega t)d\omega t + \int_{\pi+\lambda_2}^{2\pi-\lambda_2} -B d\omega t + \int_{2\pi-\lambda_2}^{2\pi} \frac{B}{a}(K_b + B_1 \sin \omega t)d\omega t\right]$$

$$= \frac{B}{\pi}\left[\frac{K_b}{a}(\lambda_1 + \lambda_2) + \frac{B_1}{a}(\cos\lambda_2 - \cos\lambda_1) + (\lambda_2 - \lambda_1)\right] \tag{21}$$

where

$$\lambda_1 = \sin^{-1}\left(\frac{a - K_b}{B_1}\right)$$

$$\lambda_2 = \sin^{-1}\left(\frac{a + K_b}{B_1}\right)$$

The quantities λ_1 and λ_2 may be expanded into a power series in $(a - K_b)/B_1$, and $(a + K_b)/B_1$ respectively. Assume that $B_1 >> K_b$ (B_1 dominates K_b), $B_1 > a + K_b$, and that the fifth or higher power terms in the series expansion can be neglected. It follows that

$$\lambda_1 = \sin^{-1}\left(\frac{a - K_b}{B_1}\right) \cong \frac{a - K_b}{B_1} + \frac{1}{6}\left(\frac{a - K_b}{B_1}\right)^3 \tag{22}$$

$$\lambda_2 = \sin^{-1}\left(\frac{a + K_b}{B_1}\right) \cong \frac{a + K_b}{B_1} + \frac{1}{6}\left(\frac{a + K_b}{B_1}\right)^3 \tag{23}$$

$$\cos\lambda_1 = 1 - \frac{a^2 - 2aK_b}{2B_1^2} - \frac{(a^2 - 2aK_b)^2}{8B_1^4} \tag{24}$$

$$\cos\lambda_2 = 1 - \frac{a^2 + 2aK_b}{2B_1^2} - \frac{(a^2 + 2aK_b)^2}{8B_1^4} \tag{25}$$

Substituting λ_1, λ_2, $\cos\lambda_1$, $\cos\lambda_2$ from these equations into equation 21, and dividing both sides by the average value K_b of the input, one obtains

$$g_b = \lim_{K_b \to 0} \frac{A_y}{K_b} = \frac{B}{a}\frac{1}{\pi}\left[2\left(\frac{a}{B_1}\right) + \frac{1}{3}\left(\frac{a}{B_1}\right)^3\right] \tag{26}$$

Appendix II. Frequency Response

The transfer function of the controller is given by

$$K_1 + j\omega K_2$$

Multiplying the complex form C_1 of the controller input, given in equation 15, by this transfer function and equating this to the complex form of the limiter input, one has

$$B_1 e^{j(\omega t + \theta)} = (K_1 + j\omega K_2) \times [K_2 + M_1 H_1 e^{j(\beta - 2\alpha)}] e^{j\omega t} \tag{27}$$

whence

$$B_1 e^{j\theta} = H_3[R_2 e^{j\beta} + M_1 H_1 e^{j(\beta + \beta - 2\alpha)}] \tag{28}$$

for

$$\beta = \tan^{-1}\omega\frac{K_2}{K_1}$$

$$H_3 = \sqrt{K_1^2 + \omega^2 K_2^2}$$

Multiplying both sides of equation 28 by $e^{-j\theta}$, one obtains

$$B_1 = H_3[R_2 e^{j(\beta - \theta)} + M_1 H_1 e^{j(\beta - 2\alpha)}] = H_3\{[R_2 \cos(\beta - \theta) + M_1 H_1 \cos(\beta - 2\alpha)] + j[R_2 \sin(\beta - \theta) + M_1 H_1 \sin(\beta - 2\alpha)]\} \tag{29}$$

Since B_1 must be real, the imaginary part of equation 29 must be zero, whence

$$\sin(\beta - \theta) = \frac{M_1 H_1}{R_2}\sin(2\alpha - \beta) \tag{30}$$

and

$$\cos(\beta - \theta) = \pm\sqrt{1 - \left[\frac{M_1 H_1}{R_2}\sin(2\alpha - \beta)\right]^2} \tag{31}$$

Substituting for $\cos(\beta - \theta)$ from equation 31 into equation 29, one obtains equation 16.

Appendix III. Equivalent Gain for Good Transient Response

To determine how the roots of equation 17 are affected by the equivalent gain g_b, the following right to left synthetic division is performed,[2] where it is assumed that $g_b = 2$.

1	20	100	$80g_b$	$100g_b$	1,	20,	100
			g_b	$20g_b$			
			100	$60g_b$		$-0.6g_b$,	g_b
			$0.6g_b$	$12g_b$			
1	20	100					

This division shows that the roots of equation 17 are approximately equal to the roots of the equation

$$x^2 + 20x + 100 = 0 \tag{32}$$

and the equation

$$x^2 + 0.6g_b x + g_b = 0 \tag{33}$$

The roots of equations 32 and 33 are

$$-10, -10, -0.3g_b \pm j\sqrt{g_b} \tag{34}$$

for practical purposes. If $g_b < 1.25$ the imaginary part of the complex roots is at

least three times the real part. The roots are then too oscillatory.

For $p_D = 1.6$ the roots are -13, -5.84, $-0.59 \pm 1.33j$; whereas for $p_D = 4$ they are -14, -2.2, $-1.75 \pm 3.2j$. For $p_D > 4$ the roots are more oscillatory and less satisfactory. For good transients the gain p_D should therefore be in the range $-1.6 < p_D < 4.0$.

References

1. Signal Stabilization of a Control System, **Rufus Oldenburger.** *ASME Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 79, no. 8, 1957, pp. 1869–72.

2. Fundamental Theory of Servomechanisms (book), **LeRoy A. MacColl.** D. Van Nostrand Company, Inc., Princeton, N. J., 1945, pp. 78–87.

3. On Asynchronous Action, **N. Minoraky.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 259, no. 3, 1955, pp. 209–19.

4. Frequency Response (book), **Rufus Oldenburger.** American Society of Mechanical Engineers; also Macmillan Company, New York, N. Y., 1956, p. 242.

5. Algebraic Approach to Design of Automatic Controls, **Rufus Oldenburger.** *ASME Transactions*, American Society of Mechanical Engineers, vol. 80, no. 2, 1958, pp. 433–43.

Discussion

Takashi Nakada (Purdue University, Lafayette, Ind.): The authors state that signal stabilization is analogous to, but differs from, the insertion of an amplifier or attenuator into the feedback loop. The difference may readily be seen by considering a loop with a dead zone nonlinear element. Such a dead zone yields a zero gain open loop. Insertion of an amplifier or attenuator ahead of this nonlinear element will not change the zero open loop gain. However, a properly chosen stabilizing signal will change the open-loop gain to a value different from zero.

R. Oldenburger and **C. C. Liu:** The example cited by Prof. Nakada emphasizes a difference between the technique of signal stabilization and that of inserting an amplifier into the feedback loop. When the open-loop gain of the unstabilized system is different from zero the results due to the insertion of a stabilizing signal are closer to those achieved with an amplifier or attenuator. The performance of a stable system can be altered within limits by the use of the "stabilizing" signal, i.e., by the introduction of an extra signal, as is done when the gain of an amplifier in the loop is changed. The paper shows how the extra signal can be determined by calculation. Replacing the nonlinear element by an amplifier (or attenuator) makes it possible to compute by conventional methods the gain this amplifier must have to yield desired performance. This is then the equivalent gain to be obtained for the nonlinear element. From the relation between the extra signal and the equivalent gain of the nonlinear element one can determine the signal that will yield the correct equivalent gain. The approach is valid for systems with multiple nonlinearities.

A Study of Nonlinear Systems With Random Inputs

KUEI CHUANG NONMEMBER AIEE

LOUIS F. KAZDA MEMBER AIEE

IT IS THE PURPOSE of this paper to present the results of investigating the probability distribution of the response of a nonlinear system when it is subjected to a Gaussian random input. The mathematical theory of a Markoff random process is used; the associated Fokker-Planck equation for a generalized continuous nonlinear system is presented. The justification for assuming the error response of a nonlinear control system as being a Markoff process is discussed.

The theoretical result of this investigation for the stationary case is obtained in a closed form. It was verified by an experimental method which used an automatic electronic counting device to record the percentage of total observation time that a random signal exceeded a certain preset voltage level. The agreement between the theoretical calculations and the experimental measurements lies within the limits of experimental accuracy.

From the results of this investigation the following conclusions may be drawn:

1. The probability density function of the error of a second-order nonlinear system is an exponential function of the form $\exp[-A \int K(\epsilon) d\epsilon]$, where $K(\epsilon)$ describes the behavior of the nonlinear element.

2. If $K(\epsilon)$ is approximately linear, then the probability density function of item one reduces to a Gaussian function of ϵ.

The response of a nonlinear system to a random input usually cannot be obtained with the use of the phase space or the describing function methods of analysis, since these methods are applicable to only a nonlinear system which contains only prescribed functions for its input. On the other hand, when a system contains a random input, this input can be characterized only by a set of statistical properties. To overcome this difficulty, Booton first introduced a linearization technique based upon the assumption that the probability distribution of the response of a nonlinear closed-loop system containing a zero memory nonlinearity, when excited by a Gaussian random input, is approximately Gaussian.[1] This statistical linearization method has been checked experimentally and the results agree with the theoretical calculations within the limits of experimental measurement.

This paper determines the theoretical probability distributions of Booton's type of nonlinear system when some special conditions are imposed on the random input function. It simultaneously explains, therefore, the validity of Booton's assumption under these special conditions.

Paper 59-147, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 22, 1958.

Kuei Chuang and Louis F. Kazda are with the University of Michigan, Ann Arbor, Mich.

This work contains part of the results of a thesis submitted by Kuei Chuang in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Michigan.

I. Statistical Idea From Phase Space Point of View

It is well known by most engineers that the response of any dynamical system can be represented by a point in phase space. For a second-order system the phase space reduces to a phase plane. In this plane a point represents the position and the velocity of the dynamical system at a certain instant of time. With increasing time the representative point, which is continuously moving in this plane, describes a phase trajectory. Its path is analogous to the stream line of flow in hydrodynamics, and it represents the history of the dynamical system. If the input to the system is a random function of time which has a wide uniform power spectrum, the representative point of the system, when described in the phase space, undergoes a random motion. This random motion, if observed on an oscilloscope, reminds one of

the Brownian movement of a particle on the surface of liquid. It is this analogy that led the authors to consider a Markoff random process technique for this motion. It is interesting to point out that the analogy between the phase space flow and the nonlinear differential equation was first considered by Kaplan.[2] The present idea of the random motion of a representative point in phase space could be considered as an extension of the ideas of phase space flow. A moment's reflection would have revealed that both of these ideas could have been derived from Boltzmann's equation.[3]

II. Conditions Imposed on Random Input Functions and Effect of Feedback

To apply the Markoff random process technique to a dynamical system, some conditions must be imposed on the random input $F(t)$, as follows:

$$\langle F(t)\rangle_{AVE} = 0 \tag{1}$$

$$\langle F(t)F(t+\tau)\rangle_{AVE} = 2D\delta(\tau) \tag{2}$$

$F(t)$ = Gaussian distributed random function (3)

Condition 2 implies that the input function $F(t)$ has a white spectrum.

By imposing conditions 1 and 2 on the random input function, one might assume that the Markoff random process technique is not useful in analyzing a feedback control system. This impression is based upon the fact that the power spectrum of the random input to a feedback control system usually is limited to a region of extremely low frequency, while condition 2 requires the random input to have a white power spectrum. However, in the following discussion it is shown that this impression is not a true one. In reality the feedback action of the system creates an effective input, which is a linear combination of the true random input and its higher-order derivatives, rather than a true random input.

Consider, for example, the first-order servomechanism which contains an integrator and a nonlinear element $K(\epsilon)$ in its forward path. The following two equations define the dynamic behavior of the system:

$$\epsilon = \theta_i - \theta_0 \tag{4}$$

and

$$\frac{d\theta_0}{dt} = K(\epsilon) \tag{5}$$

where θ_0, θ_i, and ϵ are respectively the output, input, and error. Substitution of equation 4 into equation 5 results in an equation relating the error ϵ and input θ_i:

$$\frac{d\epsilon}{dt} + K(\epsilon) = \frac{d\theta_i}{dt} = F(t) \tag{6}$$

The random function $F(t)$ is called an effective random input and is assumed to have the properties prescribed by equations 1, 2, and 3. Since $F(t)$ is a linear function of the true random input $\theta_i(t)$, which is a Gaussian random function, it follows immediately that conditions 1 and 3 are satisfied by $F(t)$. According to power spectrum analysis, the power spectrum $\phi_F(\omega)$ of the effective random input $F(t)$ and the power spectrum $\phi_{\theta_i}(\omega)$ of the true random input $\theta_i(t)$ are related by

$$\phi_F(\omega) = |j\omega|^2 \phi_{\theta_i}(\omega) \tag{7}$$

If the power spectrum $\phi_{\theta_i}(\omega)$ is of the following form:

$$\phi_{\theta_i}(\omega) = \frac{D}{(\omega^2+\alpha^2)(\omega^2+\beta^2)} \tag{8}$$

where α, β, and D are constants. Then, from equation 7 the power spectrum $\phi_F(\omega)$ of the effective random input $F(t)$ will have the following form:

$$\phi_F(\omega) = \frac{\omega^2 D}{(\omega^2+\alpha^2)(\omega^2+\beta^2)} \tag{9}$$

If in the frequency range of interest the value of the constant α^2 in equation 9 is very much smaller than ω^2, and if the value of the constant β^2 is very much larger than ω^2, the power spectrum $\phi_F(\omega)$ can be approximated by the following:

For $\omega^2 < \alpha^2$

$$\phi_F(\omega) \cong \frac{D\omega^2}{\alpha^2\beta^2} \tag{10A}$$

For $\alpha^2 < \omega^2 < \beta^2$

$$\phi_F(\omega) \cong \frac{D}{\beta^2} \tag{10B}$$

For $\omega^2 > \beta^2$

$$\phi_F(\omega) \cong \frac{D}{\omega^2} \tag{10C}$$

In Fig. 1 will be found the power spectra $\phi_{\theta_i}(\omega)$ and $\phi_F(\omega)$ as function of ω.

Inspection of Fig. 1 reveals the power spectrum of the effective random input $F(t)$ to be almost flat over the frequency range $\alpha < \omega < \beta$. If the frequency range $\alpha < \omega < \beta$ is much larger than the system bandwidth and if, in addition, the value α is very small, the power spectrum $\phi_F(\omega)$ of the effective random input may be considered as a white spectrum. (In an actual case, if a power spectrum is uniform over a frequency range which is greater than the frequency range of the system, the power spectrum may be considered as a white spectrum.) Under these conditions the effective random input has approximately the properties described by equation 2. The Markoff random process technique can be applied therefore to the first-order system provided the random input θ_i has the power spectrum $\phi_{\theta_i}(\omega)$ of the form described by equation 9.

In general, if the error of a servomechanism or of a feedback system can be described by the following equation:

$$\sum_{n=0}^{N} a_n \frac{d^n \epsilon}{dt^n} + K(\epsilon) = \sum_{n=0}^{N} a_n \frac{d^n \theta_i}{dt^n} \tag{11}$$

Since a_n's are constants, as previously defined, the right-hand side of equation 11 can be lumped together to form an effective input function $F(t)$ which is given by

$$F(t) = \sum_{n=0}^{N} a_n \frac{d^n \theta_i}{dt^n} \tag{12}$$

Accordingly, the power spectrum $\phi_F(\omega)$ of the effective input $F(t)$ and the power spectrum $\phi_{\theta_i}(\omega)$ of the true input $\theta_{i(t)}$ are related as follows:

$$\phi_F(\omega) = \left| \sum_{n=0}^{N} (j\omega)^n a_n \right|^2 \phi_{\theta_i}(\omega) \tag{13}$$

By properly choosing the form of the power spectrum $\phi_{\theta_i}(\omega)$, it is always possible to obtain a power spectrum

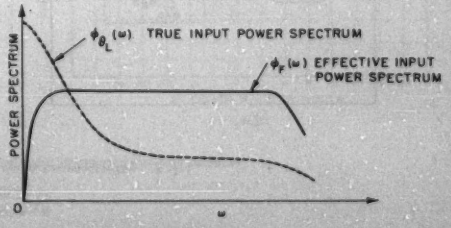


Fig. 1. Power spectrum of true and effective inputs

$\phi_F(\omega)$ which is approximately a white spectrum from a physical point of view.

The block diagram representation of equation 11 can be given in two equivalent forms: as a closed-loop representation, as shown in Fig. 2(A), and as an open-loop representation as in (B).

From Fig. 2(B) it is readily seen that the effective input to the system is essentially white noise provided that the input power spectrum is restricted to a special form determined by equation 13. Because of the power spectrum's changing property in a feedback system, it is possible to use Markoff random process technique to analyze several commonly encountered nonlinear servomechanisms.

III. Markoff Random Process and Fokker-Planck Equation

A discrete Markoff process can be interpreted as a process in which the occurrence of an event depends only on the occurrence of an event immediately preceding it. For a continuous case it is usually defined in terms of its conditional probability density function by an important relation given by equation 14, which is known as the Chapmann-Kolmogoroff equation:

$$f(t_3;\ y_1,y_2,\dots y_n|S;\ \xi_1,\xi_2,\dots\xi_n) = \int\int f(t_1;\ x_1,x_2,\dots x_n|S;\ \xi_1,\xi_2,\dots\xi_n) f(t_3;\ y_1,y_2,\dots y_n|t_1; x_1,x_2,\dots x_n)dx_1dx_2\dots dx_n \quad (14)$$

The function $f(t_1;\ x_1,\ x_2\dots x_n|\ S;\ \xi_1,\ \xi_2,\dots\ \xi_n)$ stands for the probability density function which states mathematically that at time t_1 the random variables are at $x_1,\ x_2,\dots x_n$ on the assumption that at time S they were at $\xi_1,\xi_2,\dots\ \xi_n$. For simplification in writing, vector notation is often used; equation 14 when written in vector from becomes

$$f(t_3;\ \mathbf{y}|S;\ \xi) = \int\int f(t_1;\ \mathbf{x}|S;\ \xi) f(t_3;\ \mathbf{y}|t_1;\ \mathbf{x})\mathbf{dx} \quad (15)$$

Based upon some conditions imposed on the conditional moments of those random variables, equation 15 can be reduced to a linear partial differential equation know as Fokker-Planck equation:[4]

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^{N}\frac{\partial}{\partial y_i}[\beta_i f] + \frac{1}{2}\sum_{i,j}^{N}\frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}[\beta_{ij} f] \quad (16)$$

where

$$\beta_i = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int(\Delta y_i) f(t+\Delta t;\ \mathbf{y}+\Delta\mathbf{y}|t;\mathbf{y})d(\Delta\mathbf{y})$$

$$\beta_{ij} = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int(\Delta y_i)(\Delta y_j) f(t+\Delta t;\ \mathbf{y}+\Delta\mathbf{y}|t\cdot\ \mathbf{y})d(\Delta\mathbf{y})$$

$$\beta_{i,j,k\dots 1} = 0$$

for any β whose subscripts have three or more letters.

IV. Second-Order Nonlinear Servo System and Its Relation to Fokker-Planck Equation

The dynamic behavior of a second-order servo system may be described by the second-order differential equation

$$\frac{d^2\theta_0}{dt^2} + a\frac{d\theta_0}{dt} + b\theta_0 = K(\epsilon) \quad (17)$$

where $K(\epsilon)$ is a function ϵ, a and b are constant, and θ_i and θ_0 are as previously defined.

By substituting $\epsilon = \theta_i - \theta_0$, equation 17 can be written as

$$\frac{d^2\epsilon}{dt^2} + a\frac{d\epsilon}{dt} + b\epsilon + K(\epsilon) = \frac{d^2\theta_i}{dt^2} + a\frac{d\theta_i}{dt} + b\theta_i \quad (18)$$

If the right side of equation 18 is written as a single function $F(t)$, the equation will reduce to

$$\frac{d^2\epsilon}{dt^2} + a\frac{d\epsilon}{dt} + b\epsilon + K(\epsilon) = F(t) \quad (19)$$

Since the function $F(t)$ is a linear combination of a Gaussian input function θ_i, it follows that $F(t)$ is a Gaussian random function. Furthermore by proper choice of the power spectrum of θ_i as shown in section II, $F(t)$ can be approximately treated as Gaussian white noise, and under this condition the Markoff random process technique will apply.

Now let $d\epsilon/dt = y_1$, $\epsilon = y_2$; then equation 19 becomes two first-order simultaneous differential equations in phase space $y_1 - y_2$:

$$\frac{dy_1}{dt} + ay_1 + by_2 + K(y_2) = F(t) \quad (20A)$$

$$\frac{dy_2}{dt} = y_1 \quad (20B)$$

Let t be increased to $t + \Delta t$ where Δt is very small compared with the response time of the system, but is not necessarily small compared with the function $F(t)$. Putting the two simultaneous equations 20(A) and 20(B) into the incremental form yields

$$\Delta y_1 + ay_1\Delta t + [by_2 + K(y_2)]\Delta t = \int_t^{t+\Delta t} F(t)dt \quad (21A)$$

$$\Delta y_2 = y_1\Delta t \quad (21B)$$

where $\int_t^{t+\Delta t} F(t)dt$ indicates that $F(t)$, in this small time interval Δt, is a rapidly changing function of time. By using equations 21(A) and (B), it is possible to calculate the functions β_i, $\beta_{i,j}$, β_{ijk}, ..., $\beta_{ij\dots m}$ of the Fokker-Planck equation given in section IV.

For the case of the second-order servomechanism expressed by equation 19, the Fokker-Planck equation has the following form:

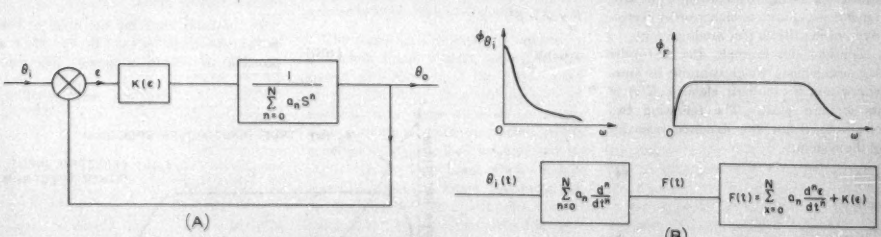


Fig. 2. Representations of nth order nonlinear servomechanism

A—Closed-loop
B—Open-loop

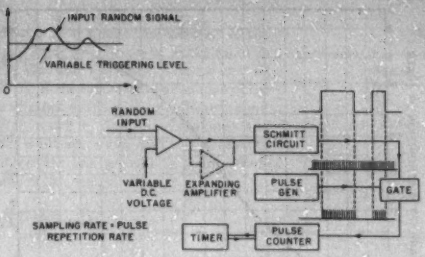


Fig. 3. Functional diagram of amplitude distribution analyzer

$$\frac{\partial f}{\partial t}=-\frac{\partial}{\partial y_2}(y_1 f)+\frac{\partial}{\partial y_1}[(a y_1+b y_2+K(y_2))f]+ D\frac{\partial^2 f}{\partial y_1^2} \quad (22)$$

Here the assumption is used that $\langle F(t)F(t+\tau)\rangle_{avg}=2D\delta(\tau)$ where $\delta(\tau)$ is a delta function and D a constant.

The solution of equation 22 gives the conditional probability density function of the nonlinear servo system. This solution, in general, is a function of y_1, y_2, t and initial conditions y_{10}, y_{20}, t_0. However, as $t\to\infty$ the solution becomes independent of y_{10}, y_{20} and t_0. This is what one would expect, since any random variable after a long time interval becomes independent of its initial conditions. Consequently, as $t\to\infty$, the conditional probability function becomes a first probability function.[5]

When the first probability density function is desired, it can be found from the steady-state solution or by letting $\partial f/\partial t=0$ in equation 22. The Fokker-Planck equation 22 reduces in this case to the following form:

$$0=-\frac{\partial}{\partial y_2}(y_1 f)+\frac{\partial}{\partial y_1}[(a y_1+b y_2+K(y_2))f]+ D\frac{\partial^2 f}{\partial y_1^2} \quad (23)$$

The additional conditions that the stationary solution must satisfy are:

1. $f(y_1,y_2,t=\infty)\to 0$ as $|y_1|\to\infty$ or as $|y_2|\to\infty$ or both $|y_1|$ and $|y_2|\to\infty$.
2. $f(y_1,y_2,t=\infty)$ should always be a positive function.
3. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(y_1,y_2,t=\infty)\,dy_1 dy_2=1$

If a stationary solution is assumed of the form $f[my_1^2+n\int[by_2+K(y_2)]dy_2]$, and these three conditions are used, it can be shown that

$$f(y_1,y_2,t=\infty)=A\exp\left\{\frac{-a}{2D}y_1^2-\frac{a}{D}\int[by_2+K(y_2)]dy_2\right\} \quad (24)$$

where A is determined by the normalized condition 3.

The steady-state solution or first probability density function in phase-space is the most general type of solution obtainable from a second-order system which has an arbitrary zero memory nonlinearity in its forward path, when the system has been subjected to a Gaussian random input. From this solution it is readily seen that in spite of a nonlinearity in this type of second-order system, the rate of the response of the

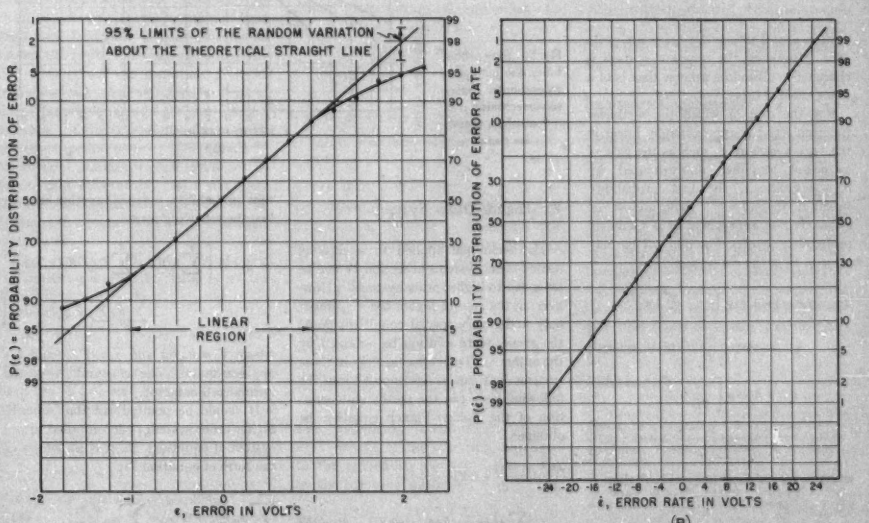

95% LIMITS OF THE RANDOM VARIATION ABOUT THE THEORETICAL STRAIGHT LINE
LINEAR REGION
P(ε) = PROBABILITY DISTRIBUTION OF ERROR
ε, ERROR IN VOLTS
(A)
P(ε̇) = PROBABILITY DISTRIBUTION OF ERROR RATE
ε̇, ERROR RATE IN VOLTS
(B)

Fig. 4. Probability distribution of second-order servomechanism with saturation

A—Error
B—Error rate

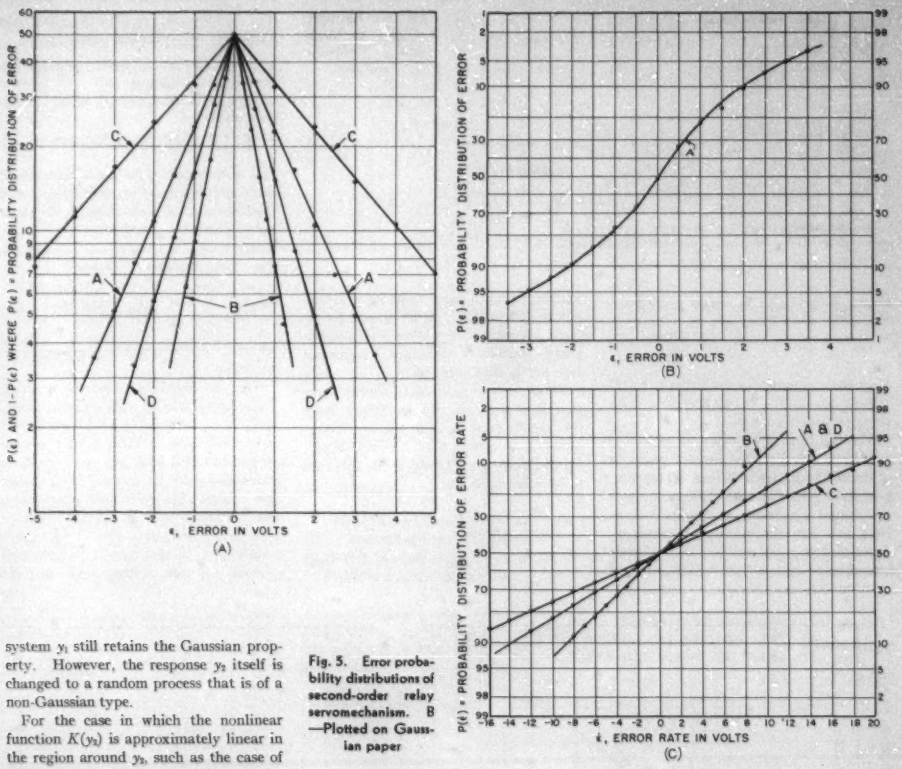

P(ε) AND 1-P(ε) WHERE P(ε) = PROBABILITY DISTRIBUTION OF ERROR
ε, ERROR IN VOLTS
(A)
P(ε) = PROBABILITY DISTRIBUTION OF ERROR
ε, ERROR IN VOLTS
(B)
P(ε) = PROBABILITY DISTRIBUTION OF ERROR RATE
ε, ERROR RATE IN VOLTS
(C)

Fig. 5. Error probability distributions of second-order relay servomechanism. B—Plotted on Gaussian paper

system y_1 still retains the Gaussian property. However, the response y_2 itself is changed to a random process that is of a non-Gaussian type.

For the case in which the nonlinear function $K(y_2)$ is approximately linear in the region around y_2, such as the case of saturated amplifier approximated by $K(y_2)=\tanh(cy_2)$, the first probability function $f(y_2,y_1,t=\infty)$ may almost be considered as approximating a Gaussian probability density function. This may be seen from the expression

$$f(y_2,y_1,t=\infty)=A\exp\left(\frac{-a^2}{2D}y_1^2\right)\times\exp\left[\frac{-a}{D}\int\tanh(cy_2)dy_2\right]$$

$$=A\exp\left[\frac{-a^2}{2D}Y_1^2\right]\times\exp\left[\frac{-a}{D}\ln\cosh(cy_2)\right]$$

when $cy_2<1$ $\ln\cosh(cy_2)\approx(cy_2)^2$ then

$$f(y_2,y_1,t=\infty)=A\exp\left[\frac{-a^2}{2D}y_1^2\right]\exp\left[\frac{-a}{D}(cy_2)^2\right]$$

which is evidently the same as a Gaussian probability density function.

V. Higher-Order System

The technique applied to a class of second-order systems in section IV can be extended to higher-order systems. However, in the case of higher-order systems, only under very special conditions could the steady-state solution be obtained by the authors. To be specific, only in those systems that can be decomposed into the following form can the steady-state solution of the Fokker-Planck equation be obtained:

$$\frac{dx_i^2}{dt^2}+a_i\frac{dx_i}{dt}+K_i(x_i)+\sum_{j=1}^{N}g_{ij}(x_j)=F_i(t),\quad i=1,2,\ldots,N \quad (25)$$

where a_i's are constants, $g_{ij}(x_j)$ and $g_{ji}(x_j)$ are of the same functional form, and $F_i(t)$'s are Gaussian white noise. The first probability density function usually has the following form:

$$f=A\exp\left[\sum_{i=1}^{N}b_iV_i^2+\sum_{i=1}^{N}C_i\int K_i(x_i)dx_i+\sum_{i=1}^{N}d_i\sum_{j=1}^{N}\int g_{ij}(x_j)dx_i\right]$$

where $V_i=dx_i/dt$ and b_i's, C_i's, and d_i's are constants. The constant A is the normalization factor.

It should be pointed out that when a higher-order system is decomposed into a system of equations 25, it is no longer of the form of equation 11:

$$\sum_{n=0}^{N}a_n\frac{d^n\theta_i}{dt^n}+K(\epsilon)=\sum_{n=0}^{N}a_n\frac{d^n\theta_i}{dt^n} \quad (11)$$

Consequently, the argument used in section III is no longer valid, because in this case one cannot write the effective

input $F(t)$ as a linear combination of the true input θ_i. However, if the input is non-Gaussian distributed and the over-all effect of the nonlinear combination of the true input θ_i still gives an effective input $F(t)$, which is Gaussian distributed and has white spectrum, then it is possible to obtain the first probability density function of this system.

VI. Computer Study

To verify the result of the analysis presented in the previous sections, an analog computer study was made on a second-order nonlinear servomechanism. Before the results of the computer study are discussed, the measuring equipment or the single-channel amplitude distribution analyzer will be described. The functional diagram of the analyzer is given in Fig. 3. This analyzer measures the probability distribution function $P(x)$ of a random variable x, the relation between the probability distribution function $P(x)$ and the probability density function $f(x)$ which is of the form:

$$P(x)=\int_{-\infty}^{x} f(x)dx \tag{26}$$

The resolution of the analyzer was checked against a known function of time such as a triangular wave. It may be shown the probability distribution function $P(x)$ versus the amplitude x of the triangular wave theoretically is a straight line. Excellent agreement was found between the theoretical straight line and the experimental points. The noise generator used was constructed in the Servomechanism Laboratory, had a flat spectrum from approximately 1/37.5 to 25 cycles per second and was Gaussian-distributed.

The equation of the second-order servomechanisms under study is of the form

$$\frac{d^2\epsilon}{dt^2}+\beta\frac{d\epsilon}{dt}+K(\epsilon)=\frac{d^2\theta_i}{dt^2}+\beta\frac{d\theta_i}{dt}=F(t) \tag{27}$$

The two different kinds of function $K(\epsilon)$ used are as follows: For the case of the saturated amplifier:

$$K(\epsilon)=\begin{cases}100, & \epsilon>1\\ 100\epsilon, & |\epsilon|<1\\ -100, & \epsilon<1\end{cases}\quad\text{and}\quad\beta=10 \tag{28}$$

For the case of the relay amplifier:

$$K(\epsilon)=\begin{cases}K=100, & \epsilon>0\\ -K=-100, & \epsilon<0\end{cases}\quad\beta=20 \tag{29}$$

In the case of equation 28 the measurements were intended to show the departure of the distribution of the error ϵ from normality in the nonlinear region; therefore, under these conditions only, is the test of normality of the error important. To confirm the belief that the departure from normality was not due to statistical fluctuation of the data, a confidence interval test was performed. The results of the experimental data, which were plotted on normal probability papers, are given in Fig. 4(A) and (B). In the second case, a detailed study was made because of the nonlinearity occurring for small values of error as well as for large. It presents, therefore, a severe test of the validity of the analysis given in previous sections.

Table I. Theoretical Calculation and Measured Data

Curve	Changed Parameter	Theoretical Calculation	Measured Data	Remarks
B	$D_N=\frac{1}{1.5^2}D_s$	$\frac{D_s}{D_N}=2.25$	$\frac{D_s}{D_N}=2.22$	D_N = changed effective input power spectrum
C	$\beta_N=\frac{1}{2}\beta_s$	$\frac{\beta_s}{\beta_N}=2.0$	$\frac{\beta_s}{\beta_N}=2.0$	β_N = changed damping
D	$K_N=1.5K_s$	$\frac{K_s}{K_N}=0.666$	$\frac{K_s}{K_N}=0.690$	K_N = changed force
A	Standard system			D_s = effective input power spectrum β_s = damping K_s = restarting force

For equation 27 with $K(\epsilon)$ and β as given in equation 29, the first probability density function $f(\epsilon,\dot{\epsilon})$, according to equation 23, is of the form

$$f(\dot{\epsilon},\epsilon)=A\exp\left(\frac{-\beta}{2D}\dot{\epsilon}^2+\frac{K\beta}{D}\epsilon\right)\quad\text{for }\epsilon<0 \tag{30A}$$

$$=A\exp\left(\frac{-\beta}{2D}\dot{\epsilon}^2-\frac{K\beta}{D}\epsilon\right)\quad\text{for }\epsilon>0 \tag{30B}$$

where A is a constant. By using the normalization condition $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(\dot{\epsilon},\epsilon)d\dot{\epsilon}\,d\epsilon=1$, it is possible to determine the constant A. The first probability of ϵ alone is determined from the relation:

$$f(\epsilon)=\int_{-\infty}^{+\infty}f(\dot{\epsilon},\epsilon)d\dot{\epsilon}$$

$$=\frac{K}{2D}\exp\left[\frac{K\beta}{D}\epsilon\right]\quad\text{for }\epsilon<0 \tag{31A}$$

$$=\frac{K}{2D}\exp\left[\frac{-K\beta}{D}\epsilon\right]\quad\text{for }\epsilon>0 \tag{31B}$$

The probability distribution function $P(\epsilon)$ defined in the following, is related to the probability density function $f(\epsilon)$ by the following relations:

$$P(\epsilon)=\int_{-\infty}^{\epsilon}f(t)dt=\frac{1}{2}\exp\left[\frac{K\beta}{D}\epsilon\right]\quad\text{for }\epsilon<0 \tag{32A}$$

$$P(\epsilon)=1-\frac{1}{2}\exp\left[\frac{-K\beta}{D}\epsilon\right]\quad\text{for }\epsilon>0 \tag{32B}$$

Since the functions $P(\epsilon)$ and $1-P(\epsilon)$ are equal to some exponential functions of ϵ, it is evident that the graph of the data plotted on semilogarithmic paper should be a straight line. Furthermore, if $\epsilon=-D/\beta K$ is substituted in equation 32(A), $P(\epsilon)$ is found to equal 0.1834. Based upon this fact, by finding the point on the graph of $P(\epsilon)$ versus ϵ at which the function $P(\epsilon)$ is equal to 18.34%, it is possible to determine the value of ϵ which corresponds to the quantity $D/\beta K$. The results of the experimental data, plotted on semilogarithmic paper and on normal probability paper, are given in Fig. 5(A), (B), and (C). The effects of the parameters on the probability distribution were checked by comparing the change with an arbitrarily chosen standard system. The theoretical calculation and the measured data are listed in Table I, and the results were amazingly close.

Summary Conclusions

The purpose of this paper is to present a method of determining the first probability density function of a nonlinear system subjected to a Gaussian random input. The Markoff random process technique used. The results from the computer study agree with the theoretical calculations within the limits of experimental error. Although this study is limited to a Gaussian input with some definite power spectrum, this analysis can be applied to a large class of nonlinear systems.

References

1. The Analysis of Nonlinear Control Systems With Random Input, R. C. Booton, Jr. *Proceedings, Symposium on Nonlinear Circuit Analysis*, Polytechnic Institute, Brooklyn, N. Y., vol. 2, 1953, p. 369.

2. Some Methods for Analysis of the Flow in Phase-Space, W. Kaplan. *Proceedings, Symposium on Nonlinear Circuit Analysis, Ibid.*, p. 98.

3. M. C. Wang. *Ph.D. Thesis*, University of Michigan, Ann Arbor, Mich.

4. M. C. Wang, G. E. Uhlenbeck. *Reviews of Modern Physics*, New York, N. Y., vol. 17, no. 2 and no. 3, 1945, pp. 323-42.

5. Analytischne Methoden der Wahrscheinlichkeitsrechnung, A. N. Kolmogoroff. *Mathematische Annaler*, Berlin, Germany, vol. 104, 1931, p. 436.

Elements of Reactor-Controlled Reversible Induction-Motor Drives

WERNER LEONHARD
ASSOCIATE MEMBER AIEE

THE PROBLEM of controlling continuously the speed of induction motors is as old as the motor itself. Numerous schemes have been developed and successfully applied since the end of the last century. Most of them are relatively efficient but require auxiliary machines. Some are inherently critical in manufacture and operation. As a result, when efficient and continuous speed control over a wide range is necessary, the adjustable-voltage d-c drive is still the most frequent choice.

In a number of applications in the low- and medium-power field, high efficiency is not of primary concern so that speed adjustment under load suffices. This is true for many intermittent drives encountered in material handling processes, overhead cranes, and the like. If induction motors are to be used, they could have secondary resistor control for motoring, and for braking overhauling load, their stators could be excited by direct current, be unsymmetrically connected, or a load brake could be added. Such drives are inefficient except at full speed, but they are simple and inexpensive. Naturally, the quality of control, i.e., the shape of the speed-torque curves obtainable, is very limited.

More recently, techniques, common with small quarter-phase servomotors, have been successfully employed for reversible intermittent drives at ratings beyond 100 kw. Typical examples are crane hoist, bridge, and trolley drives, and, at lower power, rudder control of ships or electrode control of arc furnaces.[1-6] The circuits employ static components, such as saturable reactors and magnetic amplifiers, to control the primary voltages of high-slip induction motors, usually in dependence of the motor speed. Motoring and braking torques for both directions of rotation are available without the use of reversing contactors. The speed-torque curves resemble those of Ward-Leonard drives with smooth speed control extending down to about 1% of synchronous speed. Satisfactory dynamic response and relative stability can be accomplished. Since the efficiency is low at reduced speed, as with secondary-resistor control, the application is normally restricted to intermittent slow-speed duty. The advantages of standard induction motors in conjunction with static-control apparatus are evident since the combination assures high reliability and long-life expectancy.

The historic, but apparently isolated, example of a nonreversing reactor-controlled induction-motor drive was developed by Alexanderson[7] to trim the speed of a 200-kw high-frequency alternator in his transoceanic radio-communication system.

The intent of this paper is to discuss and compare several alternative reactor circuits for reversible-motor control on the basis of their open-loop properties. No effort is made to describe an over-all speed-control system.

Principle of Operation

Symmetrical Control

The two solid curves in Fig. 1 represent the speed-torque curves of a polyphase wound-rotor motor with fixed secondary resistance, when the motor is fed with rated symmetrical line voltages, V_{mo}, of alternative sequence. Each curve consists of three branches, describing oversynchronous regenerative braking, undersynchronous motoring, and counter-torque braking operation. When varying the secondary resistance R_2, the curves are compressed or expanded in the vertical direction, since the slip at which a particular torque, for example pull-out torque, occurs is proportional to R_2. When the motor voltages, V_m, are reduced symmetrically, the curves can also be compressed in the horizontal direction by the approximate factor $(V_m/V_{mo})^2$, thus covering the area between the solid lines and the ordinate axis. Curve a in Fig. 1 is an example for $V_m/V_{mo}=0.5$. The reactor-control circuits, to be discussed, are primarily voltage-control schemes, but a variation of the secondary resistance is usually needed when a wide speed range is required. Motor efficiency varies proportionally with the motor speed, when the primary copper, iron, and friction losses are neglected:

$$\eta \lesssim n/n_0 \qquad (1)$$

This simplification is justified at high slip, where most of the losses occur in the rotor circuit. With wound-rotor motors, the rotor losses present no serious cooling problem due to their mostly being generated outside the motor. Induction motors with eddy-current-type rotors, which would dispose of the slip rings and secondary-resistor control, would require forced cooling except for very light slow-speed duty.

In Fig. 2 a static scheme for voltage control and sequence inversion is drawn.[8,9] The figure shows a 3-phase motor connected with its primary terminals into a network of six statically adjustable reactors. If their admittances are varied such that either

$$g_{min} \leqq g_1 \leqq g_{max} \qquad g_2 = g_{min} \approx 0 \qquad (2)$$

or

$$g_{min} \leqq g_2 \leqq g_{max} \qquad g_1 = g_{min} \approx 0 \qquad (3)$$

holds, the motor is supplied with adjustable voltages of alternative sequence. The two reactors connected to line terminal 3 may be combined to a single unit.

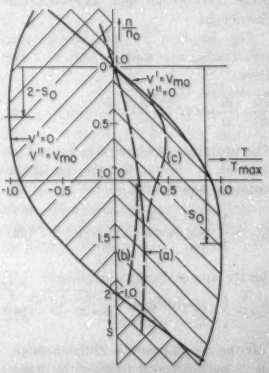


Fig. 1. Speed-torque curves of wound-rotor induction motor

Paper 58-1176, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958; and re-presented at the Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted May 26, 1958; made available for printing July 31, 1958.

Werner Leonhard, formerly with Westinghouse Electric Corporation, Pittsburgh, Pa., is now in Stuttgart, Germany.

The contributions of B. H. Beattie of the New Products Engineering Department, Westinghouse Electric Corporation, are gratefully acknowledged. The author wishes also to thank H. A. Zollinger for presenting the paper at the AIEE Winter General Meeting.

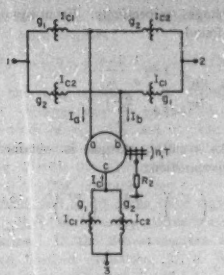


Fig. 2. Wound-rotor motor with static voltage control

The speed-torque curves for any given external-reactance setting will differ from those obtained with adjustable-voltage control. The reason is that the speed-dependent motor impedances interact with the fixed external reactances in such a manner that the motor voltages and hence the torque are reduced as the slip is increased. Fig. 1 demonstrates this effect where curve *a* for constant motor voltage is now transformed into curve *b* for constant external reactances.

The reactors in Fig. 2 usually have the form of d-c controlled saturable reactors or self-saturating magnetic amplifiers in doubler connection. Both circuits produce harmonics and both have nonlinear average voltage-current characteristics. The first complication is neglected, but the second may be qualitatively discussed. Figs. 3(A) and (B) depict the idealized output characteristics of a saturable reactor and a self-saturating magnetic amplifier. The "low-impedance" output of the self-saturating amplifier will make the motor characteristic for constant amplifier-control current approach those obtained with straight voltage control (*a* in Fig. 1). In turn, the saturable reactors with their high incremental impedance in the operating region will distort greatly the motor characteristics. At speeds close to synchronism, the motor impedance is high; correspondingly, most of the line voltage appears across the motor, while the reactors, operating in their initial low-impedance region, absorb little voltage. Hence the torque approaches that obtained with full motor voltage. When the slip increases, the motor impedance becomes lower and the currents rise. At a certain value of the motor current, determined by the reactor control current I_c, the reactor voltages increase sharply. As a result, the motor voltage decreases and the torque falls off rapidly (Fig. 1, curve *c*). This effect may introduce considerable negative slope ($\partial T/\partial s$) at speeds where the original motor characteristic still has definite positive slope. The distortion is most undesirable since it may cause unstable motor operation with constant-torque loads. Although the motor can be safely stabilized by a speed-feedback loop, its design is more difficult. The system then becomes a case of "conditional stability."

The preceding discussion and other aspects, such as reactor size, power gain, and response time, favor the self-saturating amplifiers, but the fact that they introduce rectifiers with relatively low short-time overload capacity into the power circuit limits their application. All the following discussions will deal with the saturable-reactor approach.

Asymmetrical Control

Symmetrical operation is increasingly desirable at higher power ratings where the effect of asymmetrical load on the supply must be considered. On the other hand, past experience has shown that a limited unbalance is not objectionable with high-slip motors of small and medium ratings. Since the size and the cost of the control equipment for reversible drives can be reduced when tolerating asymmetry in the motor supply, these circuits have gained some importance in practice. The restriction to high-slip motors is very important, since with low slip, such as normal National Electrical Manufacturers' Association *B* squirrel-cage motors, the customary dislike of asymmetry in the supply is fully justified. There, already a slight asymmetry, which hardly produces a control effect on the motor characteristics in the usual speed range, increases the motor currents by a substantial amount.[10,11]

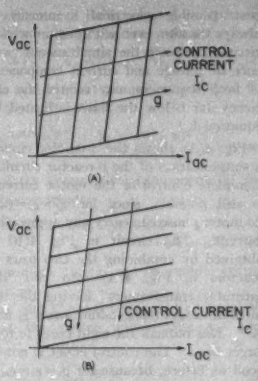


Fig. 3. Voltage-current characteristics of control components

A—Saturable reactors
B—Self-saturating magnetic amplifier

When voltage and current components, $\mathbf{V}'$, $\mathbf{I}'$ and $\mathbf{V}''$, $\mathbf{I}''$ of positive and negative sequences are simultaneously allowed at the motor terminals, the speed-torque area of Fig. 1 may be covered with a multitude of combinations between $\mathbf{V}'$, $\mathbf{V}''$, and R_s. (A zero sequence component cannot exist in the terminal voltages and currents of a symmetrically wound motor when the neutral of its star-connected stator winding is isolated or when its delta-connected stator winding is not altered.) The degree of freedom is illustrated by the fact that the starting torque of an induction motor with a given secondary resistance is proportional to the area enclosed by the vector diagram of the three sinusoidal stator voltages, independent of the shape of the vector diagram.[12] In practice, however, the

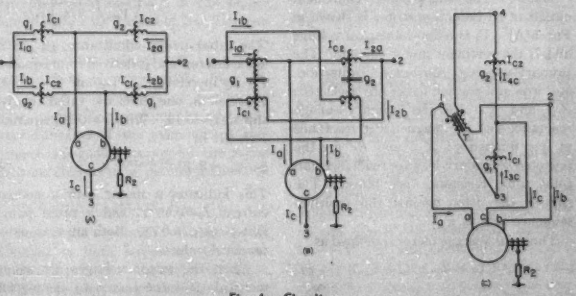


Fig. 4. Circuits

A—4-reactor circuit B—2-saturable transformer circuit C—2-reactor circuit

best possible electrical symmetry is always the aim, even with unsymmetrical circuits, because the simultaneous presence of voltage and current components of both sequences may reduce the efficiency far below the limit indicated by equation 1.

Fig. 4(A) shows the 4-reactor circuit, a simplification of the 6-reactor circuit.[13] Complete control of the motor currents is still possible, since for $g_1 = g_2 = g_{min}$, no motor phase can carry any appreciable current. The circuit in Fig. 4(B) is obtained by combining the two pairs of reactors in Fig. 4(A) into two "d-c saturable transformers" having the adjustable magnetizing admittances g_1 and g_2.[14] This reduces the eight cores to four larger ones. The control effect is not as good as before, because for $g_1 = g_2 = g_{min}$, the motor operates on single phase. The same is true for the 2-reactor circuit of Fig. 4(C) which has been frequently applied for crane controls.[2,4] Two terminals of the motor (*a,b*) are directly connected to two lines (*1,2*). The third motor terminal, *c*, can, by static control of the reactors g_1 and g_2, be positioned between the line (*3*) and the transformer terminal (*4*). Thus, the motor voltages can be made to approach symmetry with positive or negative sequence.

Approximate Analysis of Several Circuits

The circuits described in the preceding section are highly nonlinear and an exact analysis would meet with considerable difficulties. Since our present concern is not an accurate design calculation, but rather a quick comparison of the four circuits on the basis of equal assumptions, a simplified analysis will suffice.

Assumptions

Motor. The single-phase equivalent circuit of an induction motor is shown in Fig. 5(A). V_1 is a line-to-neutral voltage and I_1 the pertinent motor current. The primary-winding resistance, saturation, and iron and friction losses are neglected; X_{L1} and X_m are the primary-leakage reactance and the magnetizing reactance at line frequency; X_{L2}' and R_2' are the reflected secondary-leakage reactance and total rotor resistance per phase. The slip of the rotor against the positive-sequence field is s.

The total leakage factor is defined as

$$\sigma = 1 - X_m^2/(X_m + X_{L1})(X_m + X_{L2}') \qquad (4)$$

the pull-out slip is

$$s_0 = R_2'/\sigma(X_m + X_{L2}') \qquad (5)$$

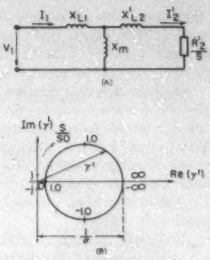


Fig. 5. Induction motor

A—Equivalent circuit
B—Normalized primary admittance

and the motor admittance at no load, $s = 0$, is

$$g_{m0} = 1/j(X_m + X_{L1}) \qquad (6)$$

With these definitions, the normalized single-phase motor admittance, g_m', may be written as

$$\gamma' = \frac{g_m'}{g_{m0}} = \frac{1 + j\frac{1}{\sigma}\frac{s}{s_0}}{1 + j\frac{s}{s_0}} \qquad (7)$$

Equation 7 is the well-known circular diagram for the primary admittance of the induction motor; Fig. 5(B). It is completely determined by the relative slip s/s_0 and the leakage factor σ, which is normally around 0.10.

With the foregoing simplifications, the torque T is proportional to the input power of the motor. Assuming a standard overload capacity $T_{max} = 2.75\ T_r$, the torque may be expressed in terms of the rated torque T_r:

$$\frac{T}{T_r} = \left(\frac{V_m}{V_{m0}}\right)^2 \frac{2T_{max}/T_r}{s/s_0 + s_0/s} = \left(\frac{V_m}{V_{m0}}\right)^2 \frac{5.5}{s/s_0 + s_0/s} \qquad (8)$$

The rated motor admittance, g_{mr}, is of some interest and is derived from equation 7. By inserting $V_m = V_{m0}$ and $T = T_r$ into equation 8, one finds the rated relative slip $s_r/s_0 = 0.19$. With $\sigma = 0.10$, equation 7 results in

$$g_{mr}/g_{m0} = 2.10\epsilon^{\pm 31}\ \text{degrees} \qquad (9)$$

This indicates a motor with a no-load current $I_0 = 0.48\ I_r$ and a rated power factor cos $\varphi_r = 0.78$. Both are reasonable practical values.

When the motor voltages are asymmetrical, it is necessary to distinguish between the motor admittances g_m' and g_m'' for positive- and negative-sequence voltages, respectively. From equation 7 is found

$$\gamma'' = \frac{g_m''}{g_{m0}} = \frac{1 + j\frac{1}{\sigma}\frac{2-s}{s_0}}{1 + j\frac{2-s}{s_0}} \qquad (10)$$

The resulting torque is obtained by superposition:

$$\frac{T}{T_r} = \left(\frac{V'}{V_{m0}}\right)^2 \frac{5.5}{\frac{s}{s_0} + \frac{s_0}{s}} - \left(\frac{V''}{V_{m0}}\right)^2 \frac{5.5}{\frac{2-s}{s_0} + \frac{s_0}{2-s}} \qquad (11)$$

where V' and V'' must be computed from the network to which the motor is connected.

Reactors. For ease of analysis, the nonlinear saturable reactors are replaced by linear reactors, having the reactive admittances g_1 and g_2 adjustable between the limits g_{min} and g_{max}. $g_{1,2}$ are normalized with the no-load admittance g_{m0} of the motor

$$\gamma_1 = g_1/g_{m0} \qquad \gamma_2 = g_2/g_{m0} \qquad (12)$$

In practice, γ is variable by a factor of approximately 200.

It is realized that due to the excessively nonlinear-reactor characteristics, the systems could not be described properly if one were to assume the existence of a unique relation $\gamma(I_c)$ between the reactor admittance γ and the control current I_c. Therefore, $\gamma_{1,2}$ are used as the primary parameters.

Six-Reactor Circuit, Fig. 2

The complete symmetry provided by the conditions 2 and 3 makes the analysis very simple. With equation 2 and $g_{min} = 0$, the voltage between two motor terminals, for example *a* and *b*, is, with equations 7 and 12:

$$V_{ab}/V_{12} = \gamma_1/(\gamma_1 + \gamma') \qquad (13A)$$

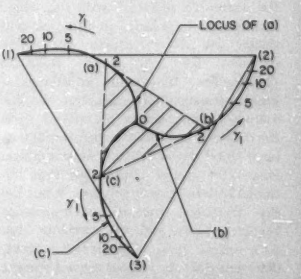


Fig. 6. Voltage diagram of 6-reactor circuit

If condition 3 is realized instead, the same voltage becomes

$$V_{ab}/V_{12} = -\gamma_2/(\gamma_1+\gamma'') \qquad (13B)$$

By inserting equation 7 or 10, the motor voltages may be calculated in dependence of $\gamma_1, s/s_0$ and σ. Fig. 6 shows the vector diagram of the voltages in Fig. 2 with the assumptions $\sigma=0.1$, $s/s_0=0.25$ (for example $s_0=4$, $s=1$, standstill) and $\gamma_2=0$. The three curves are the loci for the motor voltages as γ_1 is varied. If the other reactors are excited instead, equation 3, *a* will extend its travel towards *2* and *b* toward *1*, while *c* remains on the path between neutral and *3*.

The speed-torque characteristics for chosen values of γ_1, σ, and s_0 may be computed, for example, by inserting equation 13(A) with equation 7 into 8. The curves are very steep, similar to curve *b* in Fig. 1. Since they are not identical with the curves for constant reactor control current due to the nonlinearity of the reactors, they are not considered now. Of much greater interest, especially for the asymmetrical circuits, is the function relating the motor currents with torque and speed, since excessive motor heating is a frequent argument against reactor control. Fortunately, the nonlinearity of the reactor characteristics is of little consequence for the current-torque relation.

For the 6-reactor circuit, the motor currents are derived simply by assuming symmetrical voltage control (equations 7–9). The resulting curves in Fig. 7 show the motor current for $\sigma=0.10$ and several values of relative slip s/s_0. The end point of each curve corresponds to full motor voltage, i.e., $\gamma_1=\infty$ or $\gamma_2=\infty$. At no load, the current is zero

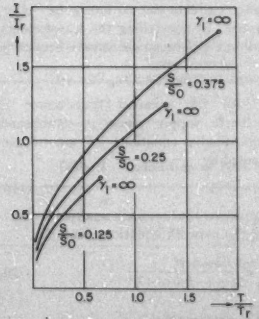


Fig. 7. Motor currents of 6-reactor circuit, symmetrical voltage control

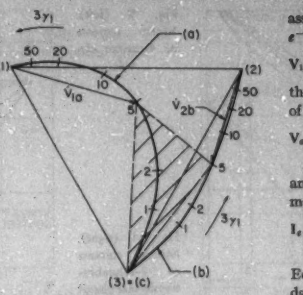


Fig. 8. Voltage diagram of 4-reactor circuit

regardless of speed, because of zero motor voltage.

In practice, the following should be noted. Symmetrical voltage control in accordance with the conditions 2 or 3, causes the torque at a given speed, to increase with the square of the motor voltage. Hence, the torque gain $\partial T/\partial\gamma$ for s=constant at light load is much lower than at full load. When the motor is controlled by a speed-feedback loop, this effects a substantial variation of the loop gain, causing the system to respond quite differently at various operating points. To overcome this, the control characteristics of g_1 and g_2 can be made overlapping such that, at low torque, both channels are partially conductive. This results in circulating currents between lines *1* and *2* and in higher motor currents, but also results in a leveling effect for the torque gain. When the motor delivers load torque of either direction, one set of reactors is driven into and remains at cut-off, while the other begins to control the motor voltages in the usual symmetrical manner. Thus, the deliberately introduced asymmetry raises the current at low load, but in the higher torque region the curves of Fig. 7 remain valid.

It has been pointed out that symmetrical control produces the best efficiency for a given set of parameters s_0, σ, s, and T; therefore, the degree of approximation of the symmetrical conditions is a good criterion for the asymmetrical circuits to be discussed now.

Four-Reactor Circuit, Fig. 4(A)

Due to the asymmetry of this circuit, voltages of both sequences are simultaneously present at the motor terminals. This is true even when following equations 2 or 3 with $g_{a,1b}=0$.

The circuit will be analyzed, with the assumption of finite g_1 and g_2. With $\alpha=e^{-j2\pi/3}$ we define the line voltages:

$$V_{12}=V_0 \qquad V_{23}=\alpha V_0 \qquad V_{31}=\alpha^2 V_0 \qquad (14)$$

the symmetrical components V' and V'' of the motor voltages:

$$V_{ab}=V'+V'' \qquad V_{bc}=\alpha V'+\alpha^2 V''$$
$$V_{ca}=\alpha^2 V'+\alpha V'' \qquad (15)$$

and the symmetrical components of the motor currents:

$$I_a=I'+I'' \qquad I_a=\alpha I'+\alpha^2 I''$$
$$I_b=\alpha^2 I'+\alpha I'' \qquad (16)$$

Equations 14, 15, and 16 contain redundancies.

I' and I'' are determined by V' and V'' and the motor admittances g_m' and g_m'':

$$I'=\frac{j}{\sqrt{3}}\,g_m' V' \qquad (17)$$

$$I''=-\frac{j}{\sqrt{3}}\,g_m'' V'' \qquad (18)$$

An inspection of Fig. 4(A) results in the following set of equations:

$$V_{ab}=V_0+\frac{1}{g_1}(I_{2b}-I_{1a}) \qquad (19)$$

$$V_{bc}=\alpha V_0-\frac{1}{g_1}I_{2b} \qquad (20)$$

$$V_0=\frac{1}{g_1}I_{1a}-\frac{1}{g_2}I_{2a} \qquad (21)$$

$$V_0=\frac{1}{g_2}I_{1b}-\frac{1}{g_1}I_{2b} \qquad (22)$$

$$I_a=I_{1a}+I_{2a} \qquad (23)$$

$$I_b=I_{1b}+I_{2b} \qquad (24)$$

Equations 14–24 are solved for V' and V''. With the abbreviations (equations 7, 10, 12), is found:

$$\frac{V'}{V_0}=\frac{3\gamma_1(\gamma_1+\gamma_2)+\gamma''(2\gamma_1+\gamma_2)}{(\gamma_1+\gamma_2)[3(\gamma_1+\gamma_2)+2(\gamma'+\gamma'')]+\gamma'\gamma''} \qquad (25)$$

$$\frac{V''}{V_0}=-\frac{3\gamma_2(\gamma_1+\gamma_2)+\gamma'(\gamma_1+2\gamma_2)}{(\gamma_1+\gamma_2)[3(\gamma_1+\gamma_2)+2(\gamma'+\gamma'')]+\gamma'\gamma''} \qquad (26)$$

As must be expected, the following symmetry conditions hold: for $\gamma_1 \longleftrightarrow \gamma_2$ and $\gamma' \longleftrightarrow \gamma''$, i.e., $s \longleftrightarrow (2-s)$, one finds: $V' \longleftrightarrow V''$ or with equation 11: $T \longleftrightarrow -T$. This confirms that equivalent operating conditions exist at any two points in the $T-n$ plane (Fig. 1), symmetrical to the origin. Equal external admittances $\gamma_1=\gamma_2$ and $\gamma'=\gamma''$, i.e., standstill, result in $V'=-V''$, or with equation 15: $V_{ab}=0$, $V_{bc}=-V_{ca}$. This indicates that the motor then operates on single-phase and, with equation 11, develops no starting torque. $\gamma_1=\gamma_2=0$ corresponds to com-

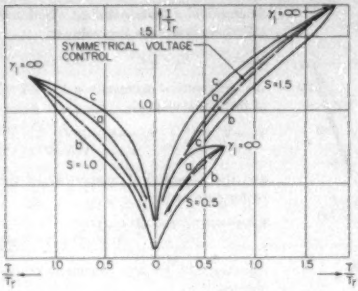


Fig. 9 (left). Motor currents of 4-reactor circuit

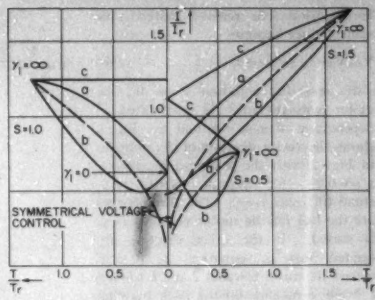


Fig. 11 (right). Motor currents of 2-saturable-transformer circuit

plete cutoff; motor voltages and torque are zero. It is recognized that due to the simultaneous occurrence of γ' and γ'' in equations 25 and 26, the specification of the relative slip s/s_0 is no longer sufficient. Both s and s_0 have to be prescribed.

Figure 8 shows the voltage diagram of the 4-reactor circuit for the following parameters: $\sigma=0.10$, $s_0=4$, $s=1$, $\gamma_2=0$ (i.e., exclusive reactor control, condition 2). When γ_1 is increased from zero, the potentials of the motor terminals a and b follow the loci towards the potentials of line terminals *1* and *2*. When g_1 is rendered nonconducting and g_2 conductive instead, the loci are interchanged, a is then traveling towards *2* and b towards *1*. Naturally, the loci are the same only if the motor is at standstill. As noted from Fig. 8, the motor voltages are definitely asymmetrical, but symmetrical conditions are approached at low and at high torque. Since the diagram is drawn for standstill, the shaded area in Fig. 8 is proportional to the starting torque. The preceding remarks with regard to uniform torque gain apply here also. Some overlap of the reactor-control characteristics is desirable to produce consistent response of the speed-feedback loop.

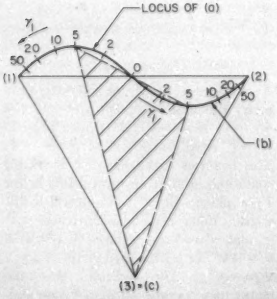


Fig. 10. Voltage diagram of 2-saturable transformer circuit

The motor current-torque function as computed from equations 7, 10, 11, 16–18, and 25–26, is again of special interest. The currents are plotted in Fig. 9 for $\sigma=0.10$, $s_0=4$, $\gamma_2=0$ and $s=0.5$, 1.0, and 1.5. For better visibility, the curves for $s=1.0$ are drawn to the left. A comparison of the currents with those obtained with symmetrical voltage control suggests that the omission of the two reactors in the third motor phase does not greatly affect the motor losses and heating. As seen later, it does allow a substantial saving in components.

Two D-C Saturable-Transformer Circuit, Fig. 4(B)

This circuit is analyzed using the same assumptions as before and equations 14–18. The normalized magnetizing admittances of the saturable transformers are γ_1 and γ_2. Their leakage reactances are neglected. The solutions for $\mathbf{V}'$ and $\mathbf{V}''$ of the network equations are

$$\mathbf{V}'/V_0=(\gamma_1+\gamma'')/(\gamma_1+\gamma_2+\gamma'+\gamma'') \quad (27)$$

$$\mathbf{V}''/V_0=-(\gamma_2+\gamma')/(\gamma_1+\gamma_2+\gamma'+\gamma'') \quad (28)$$

It is recognized that for $\gamma_1=\gamma_2=0$, i.e., no d-c excitation on either transformer, the ratio of the positive- and negative-sequence voltages equals the ratio of the two motor impedances, a characteristic of single-phase operation. Under this condition, the motor will have the well-known s-shaped speed-torque characteristic producing motoring torque in the speed range

$$|n/n_0|<\sqrt{1-(s_0\sigma)^2}$$

The single-phase torque is small for higher values of s_0. For example, $\sigma=0.10$, $s_0=4$ result in a pull-out torque of 0.16 T_r. Consequently, the negative slope $\partial T/\partial s$ in the low-speed region, introduced by single-phase operation with $\gamma_1=\gamma_2=0$, is not too disturbing.

Fig. 10 shows the vector diagram of the voltages in Fig. 4(B), again with the values $\sigma=0.1$, $s_0=4$, $s=1$, $\gamma_2=0$, and $0\leqq\gamma_1<\infty$. The diagram is drawn for standstill, hence the shaded area is proportional to the starting torque. For $\gamma_1=0$, $0\leqq\gamma_2<\infty$, the two loci for the voltages of terminals a and b are again interchanged. Due to the single-phase excitation, torque becomes immediately available when one of the control currents rises. Therefore, this circuit does not show the low gain at no load, observed with the previous circuits for exclusive reactor control (conditions 2 or 3). However, the asymmetry of the motor voltages is appreciable.

In Fig. 11, the calculated motor current-torque curves are plotted for several speeds, in comparison to those with symmetrical voltage control (values are $\sigma=0.1$, $\gamma_2=0$, $s_0=4$). The curves for $s=1$ are again drawn to the left. The no-load currents are now markedly increased. Note that at stand-by ($s=1$, $T=0$), the motor currents can no longer be "interrupted" by controlling the transformers, but that a contactor operation is necessary.

Two-Reactor Circuit, Fig. 4(C)

With the isolation transformer T, a fourth supply phase is established. Accordingly, equations 14 are supplemented by

$$\mathbf{V}_{21}=a^3V_0 \quad (14A)$$

Equations 15–18 remain unaltered. Solving the network equations leads to

$$\frac{\mathbf{V}'}{V_0}=\frac{3\gamma_1+\gamma''}{3(\gamma_1+\gamma_2)+\gamma'+\gamma''} \quad (29)$$

and

$$\frac{\mathbf{V}''}{V_0}=\frac{3\gamma_2+\gamma'}{3(\gamma_1+\gamma_2)+\gamma'+\gamma''} \quad (30)$$

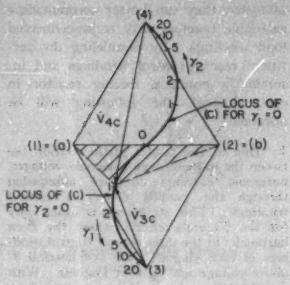


Fig. 12 (left). Voltage diagram of 2-reactor circuit

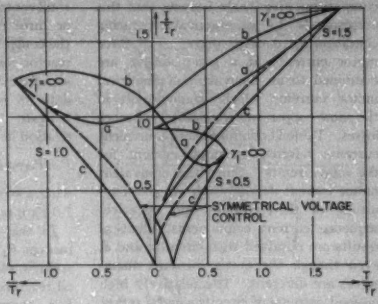


Fig. 13 (right). Motor currents of 2-reactor circuit

A comparison of equations 29 and 30 with 27 and 28 reveals that, except for the insignificant variation $\gamma_{1,x} \rightarrow 3\gamma_{1,x}$, only the sign of V^r is changed. This indicates that the same open-loop speed-torque characteristics are obtained with both circuits. The single-phase operation is, of course, much more conspicuous now, with two motor terminals directly connected to the line.

Fig. 12 represents the voltage diagram of the circuit in Fig. 4(C), assuming the previous parameters. For $\gamma_2 = 0, 0 \leq \gamma_1 < \infty$, potential c is moving toward *3*. When the other reactor is excited, c approaches *4*. Again, because of the single-phase excitation, the torque rises immediately as one reactor is excited. It is apparent that the operation of this circuit resembles most nearly, of all the schemes compared, the type of control that is common with small quarter-phase servo-motors.

In Fig. 13, the current-torque curves are plotted. They differ from those in Fig. 11 due to the different sign in equations 28 and 30. The high single-phase current at no load is again characteristic. Since only one motor current can be cut off by reactor control, a line contactor operation is necessary for stand-by. The single-phase current for $\gamma_1 = \gamma_2 = 0$, $a = 1$ may be computed as follows:

$$\frac{I_{a,b}}{I_r} = \frac{\sqrt{3}}{2}\,\frac{g_{m0}}{g_{mr}}\sqrt{\frac{1+\frac{1}{\sigma^2 s_0^2}}{1+\frac{1}{s_0^2}}} \approx \frac{\sqrt{3}}{2}\,\frac{g_{m0}}{g_{mr}}\,\frac{2}{\sigma s_0} \tag{31}$$

The approximation is valid for $s_0 \gg 1$. The ratio g_{m0}/g_{mr} of no load to rated-motor admittance is obtained from equations 7 and 8.

Discussion

Selection of s_0

Before the four circuits are compared on the basis of the calculated motor currents, the reactor sizes, and the control-power requirements, a brief remark on the selection of the secondary resistance (s_0) is appropriate. Since the correct resistance depends on the required speed range and torque, the conditions for a typical crane drive may serve as an example: the motor speed should be variable in ten steps between $s = 0$ and $s \approx 2.1$ with the limits corresponding to full-speed hoisting and lowering respectively. Between the steps, that is, mainly around standstill, $0.8 < s < 1.2$, the speed should be continuously variable. The required torque is essentially unidirectional. $T = 1.20\,T_r$ is often specified for hoisting (at $s > 0.1 - 0.13$) and $T = T_r$ for lowering speeds. A breakaway torque $1.5\,T_r$ is desirable. Reverse torque is needed during transients and in order to overcome friction, when lowering light loads.

It is evident that the speed range is too large to work with a fixed value of secondary resistance. No compromise could be struck between the contradictory requirements for sufficient motor torque when hoisting at high speed, and for acceptable motor currents and sufficient torque when lowering heavy load. Therefore, the secondary resistance must be varied in accordance with the motor speed or the speed-setting device, for example, the master switch. This complication could be avoided when using a motor with a suitable eddy-current-type rotor.

A typical set of values for the secondary resistance, expressed in terms of pull-out slip s_0 for symmetrical motor operation, is listed in Table I. It is seen that four different values for the secondary resistance are desirable. Note that for high-speed lowering, the motor is operated with negative-sequence voltages. If the hook is loaded, the motor will regenerate.

Table I. Pertinent Values of Slip and Pull-Out Slip, Symmetrical Operation for Typical Crane Drive ($s = 0$: Full-Speed Hoisting)

s	0–0.2	0.3–0.6	0.6–1.2	1.2–2.0	2.0–2.1
s_0	0.4–0.5	1.0–1.2	2.2–2.5	3.5–4	0.4–0.5

Table II. Comparison of Four Reactor Circuits; $\sigma = 0.10$, $s_0 = 4$, $s = 1.5$

	6-Reactor Circuit (a)	4-Reactor Circuit (b)	2-Saturable-Transformer Circuit (c)	2-Reactor Circuit (d)
Equivalent Symmetrical Motor Current				
$T = 0$	0	0	0.82 I_r	0.82 I_r
$T = 0.5\,T_r$	0.91 I_r	0.94 I_r	1.08	1.08
$T = T_r$	1.27	1.28	1.34	1.34
Rating Factor of Main Control Components	3.80	3.00	2.66	2.96
Relative-Control Power	1.0	0.73	0.44	0.46

Motor Currents

As an example of the motor currents to be encountered with the different control circuits, we select from the calculated current-torque curves, drawn in Figs. 7, 9, 11, and 13, those for $s_0 = 4$, $s = 1.5$. A

comparison with Table I shows that this corresponds to a practical operating condition for a crane drive. The three motor currents at a given torque are combined to an equivalent symmetrical motor current, which would result in the same copper and secondary resistance losses. Table II contains these equivalent currents referred to rated current for the four circuits. Since the equivalent motor current depends only on the magnitude of the positive- and negative-sequence current components, identical results are obtained with circuit *c* and *d*, even though the motor currents themselves are different. The relatively high no-load currents of circuits *c* and *d* reflect the large asymmetry, but as pointed out earlier, the no-load currents of circuits *a* and *b* will also rise in practice, due to overlapping reactor control. At full torque, the circuits are almost equivalent, since each approaches symmetrical conditions. This is similarly true for other suitable combinations of s and s_0 (Table I).

Rating of Control Components

As an indication of the size and cost of the major control components, a rating factor

$$f_R = \frac{\frac{1}{2}\sum V_n I_n}{\sqrt{3}\, V_m I_{mr}} \tag{32}$$

is defined relating the 2-winding transformer capacity of the saturable reactors and the isolation transformer, at the given duty cycle, to the nominal primary motor kva. The load and control windings of the saturable reactors are assumed to have equal rating. In the case of the two saturable transformers, the control windings are determined by the vectorial sum of both alternating load currents. In systems where the two reactor channels have different duty cycles, for example in hoist drives, the UP and DOWN channels may be designed for different ratings; for simplicity, it is assumed that the two channels in each circuit are equal. This is frequently so in practice. Using as a base $\sigma = 0.10$, equation 2 or 3, and the parameters $s_0 = 4$, $s = 1.5$, $T = T_r$, the rating factors of Table II were computed.

Reactor-Control Power

An important point of comparison is the amount of control power required to excite the reactors in each circuit since it affects the rating of the control source. Frequently a 2-channel magnetic amplifier or, for larger drives, a rotary exciter is used.

Since, with the different circuits, the total control power is supplied to one, two, or three reactors of different ratings P_L, the control power P_c, i.e., the losses in the reactor-control winding, must be known as a function of the reactor size. Design data for commercially available reactors indicate that the following empirical relation holds:

$$P_{c1}/P_{c2} = (P_{L1}/P_{L2})^x \tag{33}$$

with $x \approx 0.5$ for $3 < P_L < 50$ kva, and $x \approx 1.0$ for $60 < P_L < 200$ kva.

In the preceding section, the reactor ratings, P_L, were determined with $\sigma = 0.10$, $s_0 = 4$, $s = 1.5$, $T = T_r$. It is assumed that all reactors fall in the lower power range. For convenience, the control power computed for each circuit is referred to that for the 6-reactor circuit. If the relative amounts of forcing are assumed to be the same in all four circuits, the results, listed in Table II, are representative for ratings of the control sources required. Table II shows a substantial saving in control power with the last two circuits. This is to be expected because they have only one reactor in operation at a time.

Results

The rating factors, control power requirements, and consequently the costs, appear to favor the 2-saturable transformer and the 2-reactor circuit. A distinct disadvantage of these schemes, however, is the fact that the single-phase motor current cannot be reduced by reactor control. This is of special importance on crane drives having frequent stand-by periods. Also, the asymmetry of the motor voltages and currents is relatively large at low load.

The 6- and 4-reactor circuits appear more favorable in these respects. By properly biasing the magnetic control amplifiers, all motor currents can be reduced to the reactor magnetizing current, i.e., below 0.1 I_r. Thus, the motor may be kept on the line for stand-by periods whose lengths are not limited by thermal considerations. This eliminates frequent in-rush currents and voltage surges with their detrimental effects on the motor life. When comparing the 6-reactor with the 4-reactor circuit, one notes that the electrical symmetry is obtained at considerable cost. Therefore, in cases where complete symmetry is not mandatory, the 4-reactor circuit should be a good compromise.

A final argument for the 6- and 4-reactor circuits lies in the fact that they are adaptable for faster responding and more accurate drives, due to the fact that they can better accommodate saturable reactors with series-connected load windings. By paralleling the pertinent reactor-control windings and inserting a common forcing resistor in each channel, the following will be obtained:

1. Due to the proper phase relation between the induced even-harmonic voltages, harmonic currents circulate unimpeded through the parallel connected control windings. (This statement is exact only for the 6-reactor circuit where the even harmonics in the three reactor-control windings of each channel form symmetrical 3-phase voltage and current systems. With the 4-reactor circuit, it is an approximation, Fig. 8, but the circuit was found to operate very satisfactorily. The same principle was used by Alexanderson[7] in his quarter-phase motor-control scheme.) This results in lower distortion of the motor voltages.

2. The reactors do not depend on even harmonic currents flowing through the control source where they could interact with a magnetic amplifier.

3. By forcing, the reactor-response times can be effectively reduced. This is of paramount importance for obtaining fast, accurate, and stable speed-control loops.

The 2-reactor and the 2-transformer schemes have only one reactor conducting at one time. When employing series-type reactors, appreciable harmonics would appear at the motor terminals and in the control circuit having detrimental effects on the motor and a magnetic preamplifier. On the other hand, parallel-type reactors with their slow response impose serious limitations on the performance of a control system.

Comparison Tests

General

In support of the analysis, a number of laboratory tests were undertaken. A few of the results will be shown.

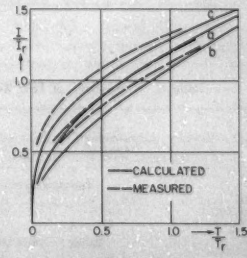


Fig. 14. Motor currents versus torque for 4-reactor circuit

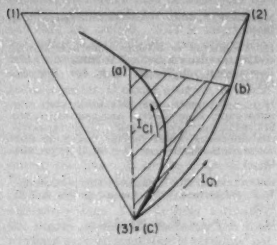


Fig. 15. Measured voltage diagram of 4-reactor circuit

$\sigma=0.11$
$I_{c2}=0$
$s_0=3.6$
$s=1.0$

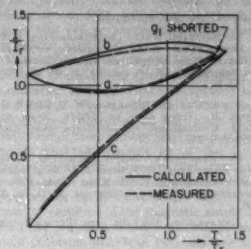


Fig. 16. Motor currents of 2-reactor circuit

$\sigma=0.11$
$\gamma_2=0$
$s_0=3.6$
$s=1.0$

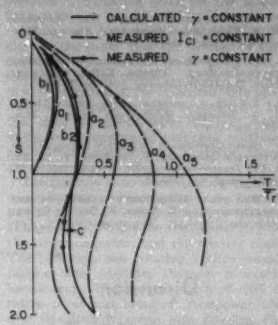


Fig. 17. Open-loop speed-torque characteristics of 2-reactor circuit

$\sigma=0.11$
$\gamma_2=0$
$s_0=3.6$

The standard wound-rotor motor used for the tests had the following specifications: 15 horsepower; 1,150 rpm; 220 volts, 42.6 amperes primary; 240 volts, 32.6 amperes secondary, cos $\varphi_r=0.79$, no-load current 19.5 amperes; maximum torque 190 foot-pounds; rated torque 69 foot-pounds.

From measurements with symmetrical voltage, a leakage factor $\sigma=0.10-0.11$ was deduced.

The 4- and the 2-reactor circuits were tested; accordingly, two types of reactors were needed:

4-reactor circuit.
A-c: 220 volts, 40 amperes
D-c: main winding, N_0, 4.35 ohms, 5.5 amperes
two auxiliary windings, 0.25 N_0, 1.30 ohms, 5.5 amperes

2-reactor circuit.
A-c: 380 volts, 40 amperes
D-c: main winding, N_0, 5.2 ohms, 5.5 amperes
two auxiliary windings, 0.25 N_0, 1.4 ohms, 5.5 amperes

The single-phase transformer for the 2-reactor circuit was designed for 220 volts, 40 amperes, primary and secondary. The ratings correspond to intermittent duty (10%). All reactors were of the parallel type; those for the 4-reactor circuit were later rewound for series operation.

Test With the 4-Reactor Circuit

As an example, the three motor currents measured with $s_0=3.6$, $s=1.5$, are compared with the results of the calculation in Fig. 14. The measurement was taken with one pair of reactors unexcited, i.e., set at minimum admittance. Considering the crude approximations in the course of the analysis, the agreement is fair. Some of the discrepancies may be explained as follows: For the calculation, it was assumed that the two conducting reactors have equal admittances, whereas in practice, they have equal control currents. If the reactor voltages differ (as seen from Fig. 8, V_{1a}, V_{2b}), the admittances will also be different. As a qualitative description, it can be stated that the constant-current branches of the reactor characteristics tend to equalize the currents in the two controlled motor phases. This is clearly seen from Fig. 14 (I_a, I_b).

Fig. 15 is the measured voltage diagram for standstill. A comparison with Fig. 8 indicates that the analysis describes the system properly.

Two-Reactor Circuit

In Fig. 16, calculated and measured motor currents are compared. The agreement is better than before because now the admittance of a single reactor is the only implicit parameter. This reduces the effect of the nonlinear-reactor characteristics.

Fig. 17 shows some open-loop speed-torque curves, measured for constant reactor control current (a_1-a_5). With both reactors set at minimum admittance, a_1 corresponds to single-phase operation. The curves b_1 and b_2 are calculated using the admittance parameters b_1: $\gamma_1=\gamma_2=0$, b_2: $\gamma_1=2/3$, $\gamma_2=0$. While b_1 is in reasonable agreement with a_1, b_2 deviates considerably from curve a_2 for constant-control current. The distortion is due to the nonlinearity of the reactors as previously discussed. This explanation is supported by curve c, which was measured with constant-reactor admittance ($\gamma_1\approx 2/3$. The curve was obtained by varying suitably the reactor control current during the test.

Conclusions

This paper presents a discussion and a simplified analysis of several reversible reactor-control schemes for intermittent induction-motor drives. The principles of the control operations are explained and numerical data are derived suitable for a comparison of the different circuits. Test results indicate that the theory, despite its simplicity, offers an adequate description of the major phenomena.

The evaluation suggests that the 4-reactor control circuit is most desirable for medium-power applications because it combines acceptable electrical symmetry with good operational characteristics (reduced line contactor duty, lack of frequent in-rush currents, fast response). At the same time, it offers a cost compromise between the simpler reactor-control schemes with their high electrical asymmetry and the fully symmetrical but more expensive 5- or 6-reactor control circuits.

The discussion is restricted to the open-loop characteristics and does not include the speed-feedback loop necessary to impart to the reactor-controlled induction motor the properties of an adjustable speed drive. Development is directed toward the more extensive use of motors with simple eddy-current rotors. By eliminating the secondary resistor control, an important step could be made towards an all static induction-motor control.

References

1. Variable-Unbalanced-Voltage Control. W. R. Wickerham. *AIEE Transactions*, vol. 64, 1945, pp. 98–102.

2. Load-O-Matic A-C Crane Hoist Control, **W. R. Wickerham.** *Westinghouse Engineer*, Pittsburgh, Pa., Mar. 1952, pp. 56–61.

3. Ein neues Elektrodenregelung mit Magnetverstärkern, **Wilhelm Kafka.** *Stahl und Eisen*, Düsseldorf, Germany, 1956, pp. 381–90.

4. A-C Crane Control with Load-O-Matic, **H. A. Zollinger.** *Westinghouse Engineer*, Mar. 1957, pp. 54–57.

5. Speed Control of Induction Motors Using Saturable Reactors, **P. L. Alger, Y. H. Ku.** *AIEE Transactions*, vol. 75, pt. III, 1956 (Feb. 1957 section), pp. 1353–41.

6. Der durch Magnetverstärker gesteuerte Asynchronmotor, **W. Kafka, W. Doell.** *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 78, 1957, pp. 884–90.

7. Transoceanic Radio Communications, **E. F. W. Alexanderson.** *AIEE Transactions*, vol. 38, pt. II, Oct. 1919, pp. 1269–85; also *General Electric Review*, Schenectady, N. Y., vol. XXIII, no. 10, 1920, pp. 794–804.

8. Motor Control Apparatus, **W. C. Carl, E. C. Rhyne, Jr.** *Patent Case no. 28,056, Serial no. 521076*, Westinghouse Electric Corporation, East Pittsburgh, Pa., 1955.

9. An Electrical Servo-Mechanism With Speed Regulation, **O. I. Aven, E. D. Demidenko, S. M. Domanitsky, E. K. Krug.** *Avtomatika i Telemekhanika*, Moscow, USSR, vol. XVII, no. 3, 1956.

10. Unbalanced Voltage and Wound-Rotor Motors, **R. F. Woll.** *Westinghouse Engineer*, May 1944.

11. Operation of 3-Phase Induction Motors on Unbalanced Voltages, **J. E. Williams.** *AIEE Transactions*, vol. 73, pt. III-A, Apr. 1954, pp. 125–33.

12. Anwendung der Methode der symmetrischen Komponenten auf unsymmetrische Ständerschaltungen von Drehstromasynchronmotoren, **Heinz Jordan.** *Archiv für Elektrotechnik*, Berlin, Germany, vol. 30, 1936, pp. 812–18.

13. Alternating Current Hoist Control, **W. R. Wickerham.** *U.S. Patent 2,440,319*, Apr. 27 1948.

14. *U.S. Patent*, **W. R. Wickerham.**

Discussion

H. Gleixner (General Electric Company, Schenectady, N. Y.): Mr. Leonhard's paper is very interesting. It describes four distinct types of reactor-controlled motor drives and it presents their advantages and disadvantages. The author suggests the 4-reactor control circuit as best from an overall viewpoint considering cost as well as performance. However, this scheme does introduce a good deal of current unbalance. As the author points out, unsymmetrical circuits cause extra motor heating and have other disadvantages.

I should like to offer for consideration a circuit that has recently been announced by the General Electric Company and has been described by L. R. Foote.[1] This circuit uses two saturable reactors and one saturable transformer. The power circuit is shown in Fig. 18. For positive torque, reactors *SX1* and *SX2* are saturated. The transformer secondary (*SXT-S*) is connected so that its output voltage is 180 degrees out of phase with its primary voltage (*SXT-P*). The motor terminals *T1*, *T2*, and *T3* are effectively connected to lines *L2*, *L1*, and *L3* respectively. For negative torque, reactor *SX2* and the transformer *SXT* are saturated. Motor terminals *T1*, *T2*, and *T3* are effectively connected to lines *L1*, *L2*, and *L3*. Intermediate values of torque are obtained by varying the amount of saturation.

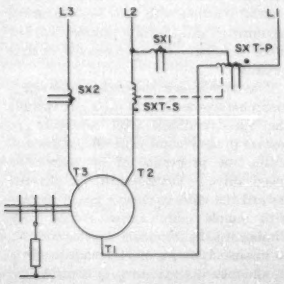


Fig. 18. Reactor-controlled reversible induction-motor drive

SX1, SX2 = saturable reactors
SXT = saturable transformer

Reactor *SX2* is used to provide the desired symmetry. The transformer primary is an autotransformer which provides compensation for the voltage drop in reactor *SX1*.

Using comparison Table II of Mr. Leonhard's paper, for the same secondary resistance values, the circuit just described should have equivalent symmetrical current values between those listed for 6-reactor and 4-reactor circuits. With reactors *SX1* and *SX2* saturated, the only unbalance is caused by the small voltage drop across the reactor *SX2*. The voltage drop across *SX1* is compensated for by the autotransformer action of the transformer. In the negative direction, a slight unbalance may occur due to some difference in impedance of the transformer and the reactor *SX2*. Nearly balanced currents are maintained throughout the entire speed range. At no time does single phasing occur.

Using a reactor-rating method which takes into account the motor current at full load, minimum slip and the maximum voltage drop across the primary control components, it is my belief that this new circuit requires considerably less main-control components rating than the 6-reactor circuit and slightly less than the 2-reactor circuit. It follows that the control-power requirements should also be less for this new circuit.

I concur with the author in that a limited unbalance is permissible for high-slip motors, particularly for wound-rotor motors with high secondary resistance. However, any unbalance causes extra motor heating, particularly in the low-slip range of operation. Therefore, the fact that the new circuit just described maintains nearly balanced currents for the full range of operation is a marked advantage.

A decrease in torque gain at light load is noted by Mr. Leonhard in paragraph 4 under "Six-Reactor Circuit, Fig. 2" in the section "Approximate Analysis of Several Circuits" describing the completely balanced 6-reactor circuit. His method of compensating for this effect is to overlap the conductance of the reversing reactors. In the case of a crane drive, it may be more desirable to retain the variation in gain in favor of the reduced line current with reduced motor and reactor heating. Further, if a stepless reference is provided for the regulator, the desired light-load speed can be obtained by manipulation of the master switch.

Under the subject of "Reactor-Control Power," the author indicates a rotary exciter for the control source of larger drives. It would be interesting to know at what horsepower Mr. Leonhard recommends changing over from magnetic amplifiers to rotating exciters.

Reference

1. Adjustable Speed Control of A-C Motors, **L. R. Foote.** *AIEE CP59-84* (available on request).

Werner Leonhard: Mr. Gleixner presents a very interesting reactor-control scheme for reversible induction-motor drives. However, some of his statements may warrant a closer examination.

In order to compare his circuit with those discussed in the paper (Table II), we refer to the same set of parameters: $\sigma = 0.10$, $s_0 = 4$, $z = 1.5$, $T = T_r$. For simplicity, we assume first that the circuit provides perfect symmetry of the motor voltages and currents ($I_m = 1.27\ I_r$). Also, we neglect the extension of winding *SXT-P* beyond the connection with *SX1*. Thus, the rating factor of the major control components is computed as: $f_R = 3.05$, which is slightly higher than the corresponding figure of the 4-reactor circuit. *SX1* and *SXT* account for the main portion of f_R. *SX1* has approximately the size of one reactor in the 2-reactor scheme (it carries the difference of two motor currents), while *SXT* is similar in size to one of the transformers in the 2-saturable transformer circuit covered in the paper.

When estimating the relative magnitude of the required control power, one finds, again with the assumptions on which Table II is based, the following figures: *SX1:* 0.48; *SX2:* 0.28; *SXT:* 0.46. For positive torque (*SX1* and *SX2* saturated), the total relative control power amounts to 0.76, and for negative torque, the figure is 0.74. These values are again very close to those of the 4-reactor circuit. When three separate amplifiers are employed for exciting the three reactors, a slight advantage in total amplifier rating but probably not in cost, is realized in that the control source for *SX2* does not need to be duplicated. When instead *SX2* is alternatively excited from one of two control amplifiers, their power rating is directly comparable to the amplifier rating needed with the 4-reactor circuit.

The new circuit is claimed to yield better symmetry than the 4-reactor circuit. Undoubtedly, measurements are given in reference 1 of Mr. Gleixner's discussion to substantiate this claim. In the meanwhile, I submit some considerations which lead me to believe that Mr. Gleixner's statements

regarding symmetry may be somewhat optimistic. Consider first the case where *SXT* and *SX2* are excited. *SX2* is inserted in one motor phase while the two others contain reactances of the two transformer load windings. If they were independent reactances as Mr. Gleixner assumes, it would seem indeed that nearly perfect symmetry could be achieved. However, since the windings are in close proximity, linking the same cores, there will be a certain amount of mutual coupling. As can be deduced from Fig. 10, any coupling between *SXT-P* and *SXT-S* will cause asymmetry of the motor voltages. When *SXT* is fully saturated, the degree of coupling between the load windings may be small but is certainly not negligible. For partial excitation of *SXT* and *SX2*, considerable coupling and hence asymmetry of the motor voltages must be expected.

Similar conditions exist when *SX1* and *SX2* are excited. The transformer *SXT*, whose primary voltage depends on the voltage across *SX1*, will cause asymmetry. Again, this will be particularly pronounced for partial excitation.

It would be interesting to know how the problem of controlling the reactors and the transformer from the control amplifiers has been solved. As pointed out in the paper, proportional control of an individual saturable reactor from a magnetic amplifier may require that the reactor be of the slow-responding-parallel type or that the harmonics in the control winding of the series-type reactor be by some means kept away from the rectifiers of the magnetic amplifier. With the proposed circuit, this will present some problems since the two reactors and the transformer are all of different ratings and do not permit convenient interconnections of the control windings as is possible with the 4-reactor circuit. There, the control windings of pairs of four equal-series-type reactors can be paralleled forming low-impedance circulating paths for the larger harmonics.

It may not have been made sufficiently clear in the paper that the overlap of the conductance regions of the reversing reactors is an optional measure. If, in a certain case, the increase of the motor currents prohibits an overlap, it may be omitted. In that case, the speed-control loop will show in the low-torque region a sluggish response and increased regulation. Otherwise, there are no detrimental effects.

The question as to when rotating exciters may become advantageous as control sources for the reactors is difficult to answer. It depends, of course, on the availability of the larger amplifiers and the relative cost. The fact that one rotating exciter would replace both amplifier channels is worth mentioning. If drives of this type with a rating of several hundred horsepower are ever designed, it seems that rotating exciters should certainly be considered.

The Linear Least Squares Synthesis of Multivariable Control Systems

R. C. AMARA
NONMEMBER AIEE

A MULTIVARIABLE system is characterized by a multiplicity of inputs and outputs. In the following, it will be assumed that such systems can be described by linear, time-in variant, input-output relations over the whole range of operation or over the restricted range of interest. Thus, let

$$R=\begin{vmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ r_p \end{vmatrix} \quad C=\begin{vmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ c_q \end{vmatrix} \quad A=\begin{vmatrix} a_{11}a_{12}\ldots a_{1p} \\ \cdot \\ \cdot \\ \cdot \\ a_{q1}\ldots a_{qp} \end{vmatrix} \tag{1}$$

then, the system input-output relation may be expressed compactly as

$$C=AR \tag{2}$$

where R, C, and A are input, output, and transfer-function matrices respectively; see Fig. 1. It is seen that a multivariable system derives each output by a simultaneous operation on all its inputs. Furthermore, it is apparent that the class of multivariable systems includes the common single-variable system as a special case.

Numerous practical examples of multivariable control systems can be cited. One of the first to receive attention as such was the turbojet engine. In this case, there are two principal inputs, namely, rate of fuel flow and exhaust nozzle area, and the two outputs, engine speed and engine temperature. Adjustment of either input affects both quantities simultaneously. A steel-rolling mill represents an example of a system with many more variables. Here the input variables can be such quantities as steel input gage, rolling speed, distance between rollers, and roller torque, while the output quantities may be steel output gage and steel tension between rolling stands. Other examples are provided by any number of manufacturing and chemical processes in which the control of each output quantity depends on the precise control of several input variables.

The general objective to be achieved by a multivariable control system may be stated simply as follows: to operate on the input variables in such a way that the output variables behave as closely as possible in a desired or prescribed manner. As in single-variable systems, feedback may or may not be used for this purpose. In addition, as in the single-variable case, a multivariable system may be assumed to operate entirely with continuous signals, entirely with discrete signals, or with combinations of both. In this connection, it may be noted that one of the principal reasons for the rapid growth in interest in multivariable systems may be traced to the possibility of using discrete elements, or digital computers, to perform the required control function.

The primary purpose of this paper is to describe a practical synthesis procedure for multivariable control systems subject to stationary stochastic inputs. Although the analysis deals only with a consideration of the continuous signal case, an extension of the method to hybrid (continuous-discrete) and discrete signal systems way be readily made.

Paper 59-214, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 27, 1958; made available for printing December 19, 1958.

R. C. AMARA is with the Stanford Research Institute, Menlo Park, Calif.

Appreciation is expressed for the continuous encouragement and advice given by Prof. G. F. Franklin of Stanford University under whose supervision this study was undertaken. Acknowledgement is also made of the stimulating guidance provided by Dr. T. H. Meisling of Stanford Research Institute.

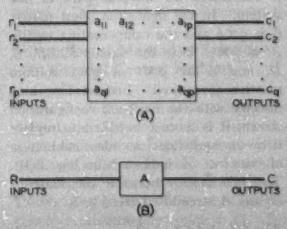


Fig. 1. Representations of linear multivariable system

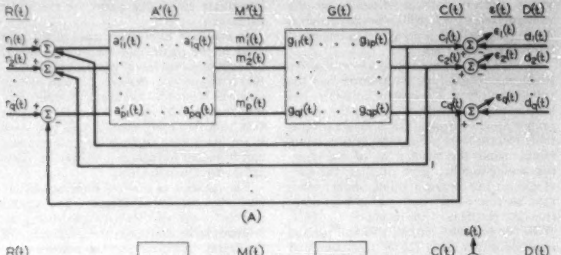


Fig. 2. Feedback and cascade representations of the general control problem

Basic System Equations

The principal properties of the systems to be considered and the criterion which is to be used as a measure of performance are stated in the following:

1. All signals are members of stationary, random, continuous processes with power spectra expressible as rational functions. In any practical problem, the usual procedure is to approximate all power spectra by rational functions in order to obtain transfer functions in useful form.

2. The fixed plant matrices as well as the operator matrices to be synthesized are linear, continuous, time-invariant, and stable; they may be represented by transfer function matrices whose elements are expressible as rational functions of complex frequency.

3. The criterion of system performance is the minimization of the sum of the mean-square errors between the set of actual outputs and a set of ideal or desired outputs. This has the merit both of lending itself to analytical treatment and of being applicable to many real problems.

The foregoing conditions or assumptions are essentially direct analogs of those normally applying to the equivalent single-variable case.

Fig. 2(A) shows the basic multivariable control system; as may be seen, direct negative feedback from output to input is considered an integral part of the system. In the figure the plant matrix G is $q\times p$ and the compensating matrix A' is $p\times q$; also, the signals R, M', C, D, and ε are $p\times 1$ or $q\times 1$ column matrices as required. Instead of working directly with the feedback configuration shown, it is more convenient to replace it by an equivalent cascade combination of matrices as indicated in Fig. 2(B). The relationship between the matrices A' and A is readily derived to be

$$A'=A(I-GA)^{-1} \tag{3}$$

Furthermore, it can be shown that no design freedom is lost in replacing one configuration by the other provided that G is stable, as assumed.

From Fig. 2(B) the total mean-square error $\overline{\varepsilon_T{}^2(t)}$ for a plant with p inputs and q outputs is

$$\overline{\varepsilon_T{}^2(t)}=\sum_{h=1}^{q}\overline{c_h{}^2(t)}-2\overline{c_h(t)d_h(t)}+\overline{d_h{}^2(t)} \tag{4}$$

where the c_h and d_h are the actual and desired outputs respectively. A direct expansion of equation 4 in terms of the elements of matrix A, the elements of matrix G, and suitably defined correlation functions yields

$$\begin{aligned}\overline{\varepsilon_T{}^2(t)}=\sum_{h=1}^{q}\Bigg[&\sum_{i=1}^{p}\sum_{j=1}^{q}\int_{-\infty}^{\infty}d\alpha g_{hi}(\alpha)\times\\&\int_{-\infty}^{\infty}d\beta g_{hk}(\beta)\int_{-\infty}^{\infty}d\sigma a_{ij}(\sigma)\times\\&\int_{-\infty}^{\infty}d\rho a_{kl}(\rho)\phi_{r_jr_l}(\sigma+\alpha-\rho-\beta)-\\&2\sum_{i=1}^{p}\sum_{j=1}^{q}\int_{-\infty}^{\infty}d\alpha g_{hi}(\alpha)\int_{-\infty}^{\infty}d\\&\sigma a_{ij}(\sigma)\phi_{r_jd_h}(\sigma+\alpha)+\phi_{d_hd_h}(0)\Bigg]\end{aligned} \tag{5}$$

where the correlation function $\phi_{xy}(\tau)$ is defined in the usual way as

$$\phi_{xy}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}dt x(t)y(t+\tau)=\overline{x(t)y(t+\tau)} \tag{6}$$

Now, a minimization of equation 5 may be performed with respect to the a_{ij} or a_{kl} by a standard variational procedure. The result is that

$$\begin{aligned}&\sum_{h=1}^{q}\sum_{i=1}^{p}\sum_{j=1}^{q}\int_{-\infty}^{\infty}d\alpha g_{hi}(\alpha)\int_{-\infty}^{\infty}d\beta g_{hk}(\beta)\times\\&\int_{-\infty}^{\infty}d\rho a_{kl}(\rho)\phi_{r_jr_l}(\sigma+\alpha-\rho-\beta)-\\&\sum_{h=1}^{q}\int_{-\infty}^{\infty}d\alpha g_{hi}(\alpha)\phi_{r_jd_h}(\sigma+\alpha)=0;\ \sigma\gtrless 0\end{aligned}$$

$$\begin{aligned}i&=1,2,\dots p\\j&=1,2,\dots q\end{aligned} \tag{7}$$

are the qp minimizing equations which must be satisfied. When the a_{kl} satisfy equation 7, the minimum realizable error $\overline{\varepsilon_T{}^2(t)}_{\min}$ is achieved. An expression for evaluating it may be obtained by noting that equation 5 reduces to

$$\begin{aligned}\overline{\varepsilon_T{}^2(t)}_{\min}=\sum_{h=1}^{q}\sum_{i=1}^{p}\sum_{j=1}^{q}\Big[&\phi_{d_hd_h}(0)-\\&\int_{-\infty}^{\infty}d\alpha g_{hi}(\alpha)\int_{-\infty}^{\infty}d\sigma a_{ij}(\sigma)\phi_{r_jd_h}\times\\&(\sigma+\alpha)\Big]\end{aligned} \tag{8}$$

when equation 7 is satisfied.

The frequency domain versions of equations 7 and 8 are most useful in obtaining numerical solutions. Equation 7 may be considered a function of σ, designated $f_{ji}(\sigma)$; when a Fourier transformation is made, the result is

$$\begin{aligned}f_{ji}(s)=&\sum_{h=1}^{q}\sum_{k=1}^{p}\sum_{l=1}^{q}g_{hk}(-s)g_{hk}(s)a_{kl}(s)\times\\&\phi_{r_jr_l}(s)-\sum_{h=1}^{q}g_{hi}(-s)\phi_{r_jd_h}(s)\\&\begin{aligned}i&=1,2,\dots p\\j&=1,2,\dots q\end{aligned}\end{aligned} \tag{9}$$

where $s=j\omega$, and $f_{ji}(s)$ has no poles in the left half of the s-plane. Since the factors of the terms within the summations are expressible as rational functions of s, a partial fraction expansion with respect to left-half s-plane poles must equate to zero.

Therefore

$$\begin{aligned}&\left[\sum_{h=1}^{q}\sum_{k=1}^{p}\sum_{l=1}^{q}g_{hk}(-s)g_{hk}(s)a_{kl}(s)\phi_{r_jr_l}(s)\right]^{+}\\&\quad=\left[\sum_{h=1}^{q}g_{hi}(-s)\phi_{r_jd_h}(s)\right]^{+}\end{aligned} \tag{10}$$

where $[\dots]^{+}$ indicates the expansion in terms of left-half s-plane poles only. Now, by using matrix notation the expression may be simplified even further. Let Φ_{rr} be the $q\times q$ matrix of elements $\phi_{r_jr_l}(s)$, Φ_{rd} be the $q\times q$ matrix of elements $\phi_{r_jd_h}(s)$, G be the $q\times p$ matrix of elements $g_{hk}(s)$, and A be the $p\times q$ matrix of elements $a_{kl}(s)$.

Equation 10 becomes

$$[\Phi_{rr}A^{T}G^{T}\bar{G}]^{+}=[\Phi_{rd}\bar{G}]^{+} \tag{11}$$

where, for any matrix $X(s)$, $X(-s)$ is written as $\bar{X}$, and X^{T} is the transpose of X. Next, equation 8 may be written in its transformed version as

$$\begin{aligned}\overline{\varepsilon_T{}^2(t)}_{\min}=&\sum_{h=1}^{q}\sum_{i=1}^{p}\sum_{l=1}^{q}\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\times\\&[\phi_{d_hd_h}(s)-g_{hi}(-s)a_{il}(-s)\phi_{r_jd_h}(s)]ds\end{aligned} \tag{12}$$

Now, letting Φ_{dd} be the $q\times q$ matrix of elements $\phi_{d_hd_g}(s)$ and using the other

matrices as defined previously yields

$$\overline{\epsilon_T{}^2(t)}_{min} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} tr\left\{\Phi_{dd} - \bar{G}\bar{A}\Phi_{rd}\right\} ds \quad (13)$$

where $tr\{\ldots\}$ denotes the trace of a matrix.

Method of Solution: Free-Configuration Control Problem

When the plant matrix G becomes the identity matrix, equation 11 reduces to the form in which the plant imposes no restrictions on the $\overline{\epsilon_T{}^2(t)}_{min}$ which may be obtained. This is sometimes referred to as a "free-configuration" design; for this case equation 11 reduces to

$$[\Phi_{rr}A^T]^+ = [\Phi_{rd}]^+ \quad (14)$$

It will be shown that the method of solution which is applied to this simpler special case can also be used in the most general control case. Accordingly, equation 14 will be used as the basis for developing the method.

Explicit solutions for matrix A^T or A may be obtained from equation 14 when matrix Φ_{rr} satisfies prescribed conditions. For example, when Φ_{rr} is diagonal, the methods of spectrum factorization developed by Wiener for the single-variable case may be applied to yield.

$$A^T = (\Phi_{rr}{}^+)^{-1}[\Phi_{rd}(\Phi_{rr}{}^-)^{-1}]^+ \quad (15)$$

In equation 13, $\Phi_{rr} = \Phi_{rr}{}^+\Phi_{rr}{}^-$, where the elements of $\Phi_{rr}{}^+$ have all their poles and zeros in the left-half s-plane and the elements of $\Phi_{rr}{}^-$ have all their poles and zeros in the right-half s-plane. When Φ_{rr} is not diagonal, matrix transformation methods may still be applicable to obtain explicit solutions. However, for the most general form of Φ_{rr} no method is known for obtaining explicit solutions for matrix A. For this case, an implicit method has been developed.

The basis for the method stems from two principal observations: 1. rational self-power and cross-power spectra result in a transfer function matrix A whose elements are also rational functions, and 2. the number of independent equations to be satisfied as a result of the partial fraction expansions of equation 14 grows more rapidly than the number of available unknowns unless the poles of each element of matrix A are initially assumed to be identical with those of all other elements of the matrix. Equipped with these observations, the method of solution may be readily derived.

For the kth row of A, equation 14 may be written in the form

$$\begin{aligned}
&[a_{k1}(s)\phi_{r_1r_1}(s)]^+ + [a_{k2}(s)\phi_{r_1r_2}(s)]^+ + \ldots \\
&\quad [a_{kq}(s)\phi_{r_1r_q}(s)]^+ = [\phi_{r_1d_k}(s)]^+ \\
&\quad\quad \vdots \qquad\qquad \vdots \\
&[a_{k1}(s)\phi_{r_qr_1}(s)]^+ + [a_{k2}(s)\phi_{r_qr_2}(s)]^+ + \ldots \\
&\quad [a_{kq}(s)\phi_{r_qr_q}(s)]^+ = [\phi_{r_qd_k}(s)]^+
\end{aligned} \quad (16)$$

If the $a_{k1}, a_{k2}, \ldots a_{kq}$ have identical poles and, in the most general case, none of these correspond to the left-half s-plane poles of the $\phi_{r_jr_1}(s)$ or $\phi_{r_jd_k}(s)$, it follows that for n such poles $-\alpha_u (u = 1,2,\ldots n)$

$$\begin{aligned}
&(s+\alpha_u)a_{k1}(s)\phi_{r_1r_1}(s)\big|_{s=-\alpha_u} + \\
&\quad (s+\alpha_u)a_{k2}(s)\phi_{r_1r_2}(s)\big|_{s=-\alpha_u} + \ldots \\
&\quad\quad (s+\alpha_u)a_{kq}(s)\phi_{r_1r_q}(s)\big|_{s=-\alpha_u} = 0 \\
&\quad\quad\quad \vdots \qquad\qquad\qquad \vdots \\
&(s+\alpha_u)a_{k1}(s)\phi_{r_qr_1}(s)\big|_{s=-\alpha_u} + \\
&\quad (s+\alpha_u)a_{k2}(s)\phi_{r_qr_2}(s)\big|_{s=-\alpha_u} + \ldots \\
&\quad\quad (s+\alpha_u)a_{kq}(s)\phi_{r_qr_q}(s)\big|_{s=-\alpha_u} = 0
\end{aligned} \quad (17)$$

It is at once apparent from the form of the equations that the matrix $\Phi_{rr}(-\alpha_u)$ can be considered the coefficient matrix of the residues of the $a_{k1}(s)$ at each pole $s = -\alpha_u$. If a nontrivial solution is to exist for the foregoing set of homogeneous equations, then the determinant of the coefficient matrix must be zero, or

$$|\Phi_{rr}(-\alpha_u)| = 0 \quad (18)$$

Since Φ_{rr} is Hermitian, equation 18 will always be an even equation in the unknowns α_u. This in turn assures that exactly half of its roots will have positive real parts; these represent poles in the left-half s-plane as required. Since Φ_{rr} is positive semidefinite as well as Hermitian, this applies even when some of the roots of equation 18 are purely imaginary; otherwise, if such roots did not occur in pairs, there would be a shift of sign along the imaginary axis, which would violate the condition of positive semidefiniteness. Thus, a division can always be made to separate the roots into two groups, one on each side of the imaginary axis. In this way, the $\alpha_1, \alpha_2, \ldots \alpha_n$ poles of the $a_{k1}(s)$ are determined, where $2n$ is the order of the determinant of Φ_{rr}.

In a manner which is somewhat analogous to that of the solution of differential equations, the foregoing poles may be considered to correspond to the natural or "force-free" poles of the system. To these poles must now be added the left-half s-plane poles of the $\phi_{r_jd_k}(s)$ (acting as forcing terms) which do not already appear as poles of the $\phi_{r_jr_1}(s)$. This completes the determination of all the poles of the $a_{k1}(s)$.

As a final step, the number of independent equations which result from the partial fraction expansions may now be determined. The number of such equations determines the number of unknowns that are available as coefficients for the assumed numerator polynomials in s of the various $a_{k1}(s)$. The distribution of these coefficients among the various $a_{k1}(s)$ can readily be determined by inspection of the power spectra of the inputs. Now, by equating residues of all poles on each side of the group of equations represented by equation 16, a set of linear algebraic equations in the unknowns results for which a solution uniquely determines the coefficients of the numerator polynomials of the $a_{k1}(s)$, or the location of the zeros. In this manner the elements of A are completely determined.

Since the preceding derivation is based on the most general form of solution, any number of special cases within this formulation are also included. For example, if the partial fraction expansions result in powers of s degrees or higher, it can be shown that the preceding method is still valid. Or if any of the natural poles of the $a_{k1}(s)$ correspond to left-half s-plane poles of the $\phi_{r_jd_k}(s)$, the solution proceeds in exactly the same manner. Similarly, it is also now apparent that the natural poles of the $a_{k1}(s)$ cannot, in general, be identical to those of any individual $\phi_{r_jr_1}(s)$; this is because the poles of the $a_{k1}(s)$ represent zeros of the determinant of Φ_{rr}, whereas the poles of the $\phi_{r_jr_1}(s)$ represent poles of the same determinant.

In recapitulation, the principal steps of the implicit method of solution are listed as follows:

1. The determinant of Φ_{rr} is set equal to zero; the left-half s-plane roots of this even polynomial in s (or α_u) determine the nat ural poles of all $a_{k1}(s)$.
2. The left-half s-plane poles of the $\phi_{r_jd_k}(s)$ not appearing in the $\phi_{r_jr_1}(s)$ are added to the poles determined in the foregoing.
3. A partial fraction expansion is made and the number of independent equations is counted. This determines the total number of polynomial coefficients for the numerators of the $a_{k1}(s)$.
4. The residues for each pole on both sides of the equations are equated and the resulting set of linear algebraic equations is solved. This completely specifies all $a_{k1}(s)$.

Adaptation of Method of Solution to General Control Problem

The implicit method of solution clearly depends on the existence of a Hermitian coefficient matrix for the elements of the unknown matrix. It will now be shown that such an "effective coefficient matrix" always does exist for the most general case as represented by equation 11. For this purpose it is more suitable to transpose both sides of equation 11 to give

$$[\Phi_{gg}{}^{T} A \Phi_{rr}{}^{T}]^{+} = [G^{T}\Phi_{rd}{}^{T}]^{+} \tag{19}$$

where the matrix Φ_{gg} has been defined by $\Phi_{gg} = G^{T}\bar{G}$. Both $\Phi_{gg}{}^{T}$ and $\Phi_{rr}{}^{T}$ are, of course, still Hermitian. If $\Phi_{gg}{}^{T}$ alone operated on A, the columns of A would have $\Phi_{gg}{}^{T}$ as their coefficient matrix; similarly, if $\Phi_{rr}{}^{T}$ alone operated on A, Φ_{rr} would be the coefficient matrix for the rows of A. Therefore it is clear that both matrices operating on A simultaneously produce a composite coefficient matrix which operates on both the rows and columns of A. The nature of this matrix will now be examined.

Let A, $\Phi_{rr}{}^{T}$ and $\Phi_{gg}{}^{T}$ be represented by

$$A = \begin{bmatrix} \rho_1{}^{A}(s) \\ \rho_2{}^{A}(s) \\ \vdots \\ \rho_p{}^{A}(s) \end{bmatrix}, \quad \Phi_{rr}{}^{T} = [\rho_1{}^{rr}(s)\rho_1{}^{rr}(s) \ldots \rho_q{}^{rr}(s)],$$

$$\Phi_{gg}{}^{T} = \begin{bmatrix} \phi_{g_1g_1}(s)\phi_{g_1g_2}(s) \ldots \phi_{g_1g_p}(s) \\ \vdots \\ \phi_{g_pg_1}(s) \ldots\ldots \phi_{g_pg_p}(s) \end{bmatrix} \tag{20}$$

where $\rho_k{}^{A}(s)$ is the kth row vector of A, $\rho_j{}^{rr}(s)$ is jth row vector of Φ_{rr}, and $\phi_{g_ug_v}(s)$ is the element in the uth row and vth column of Φ_{gg}. Therefore

$$A\Phi_{rr}{}^{T} = \begin{bmatrix} (\rho_1{}^{A} \times \rho_1{}^{rr})(\rho_1{}^{A} \times \rho_2{}^{rr}) \ldots (\rho_1{}^{A} \times \rho_q{}^{rr}) \\ \vdots \\ (\rho_p{}^{A} \times \rho_1{}^{rr}) \ldots\ldots\ldots (\rho_p{}^{A} \times \rho_q{}^{rr}) \end{bmatrix} \tag{21}$$

An examination of each row of this new matrix shows that Φ_{rr} is the coefficient matrix for each row of A. Now, the complete product is written as

$$\Phi_{gg}{}^{T} A \Phi_{rr}{}^{T} = \begin{bmatrix} \phi_{g_1g_1}\phi_{g_1g_2} \ldots\ldots \phi_{g_1g_p} \\ \vdots \\ \phi_{g_1g_p} \ldots\ldots\ldots \phi_{g_pg_p} \end{bmatrix} \begin{bmatrix} (\rho_1{}^{A} \times \rho_1{}^{rr})(\rho_1{}^{A} \times \rho_2{}^{rr}) \ldots (\rho_1{}^{A} \times \rho_q{}^{rr}) \\ \vdots \\ (\rho_p{}^{A} \times \rho_1{}^{rr}) \ldots\ldots\ldots (\rho_p{}^{A} \times \rho_q{}^{rr}) \end{bmatrix} \tag{22}$$

It is possible to deduce sufficient information about the structure of the new matrix without completely performing the foregoing operation. Each element of the resulting matrix contains all rows of A and therefore all elements of A. If a matrix composed of the rows of A taken sequentially is considered, a $pq \times 1$ vector can be formed for which an effective coefficient matrix exists. To form such a coefficient matrix, each element of the matrix of equation 22 is considered a row in a new $pq \times pq$ matrix; furthermore, each row in this new matrix is arranged in the same sequence as the corresponding element from which it was generated by standard matrix operations.

When the rows of the new matrix are examined, it is at once apparent that the vth set of q rows incorporates the factor $\phi_{g_vg_1}(s)$ in its first q columns, $\phi_{g_vg_2}(s)$ in its second set of q columns, and the factor $\phi_{g_vg_p}(s)$ in the pth set of q columns. Thus in the new matrix each element of $\Phi_{gg}{}^{T}$ exerts an influence in a $q \times q$ submatrix corresponding to its original position in $\Phi_{gg}{}^{T}$. Now, if a $pq \times pq$ matrix is constructed in which each of its submatrices consists of a $q \times q$ matrix of identical elements of $\Phi_{gg}{}^{T}$, the first portion of the effective coefficient matrix would be thus formed. It is easily verified by inspection that such a matrix, $\mathring{\Phi}_{gg}$ is always Hermitian; see Fig. 3. To determine the contribution of Φ_{rr} to the coefficient matrix, it is seen that performing the operations indicated by equation 22 involves sequencing from column to column as each element is generated. Since each such element is a row of the new matrix, each q successive rows of this matrix contain the q rows of Φ_{rr} in sequence. Thus, in the new matrix, designated $\mathring{\Phi}_{rr}$, Φ_{rr}, appears intact as a $q \times q$ matrix repeated $p \times p$ times. That this second portion of the effective coefficient matrix is also Hermitian can also be easily verified by inspection; see Fig. 3. The composite matrix may thus now be constructed by an operation of matrix superposition. This operation is defined as follows: Two matrices X and Y of the same dimensions are superposed by multiplying corresponding elements of each matrix to obtain the corresponding element of the new matrix, $W = (X \diamondsuit Y)$. In Fig. 3 the resulting matrix is the $pq \times pq$ coefficient matrix K; it is also Hermitian since it is clear that superposition preserves the Hermitian property of the superposed matrices.

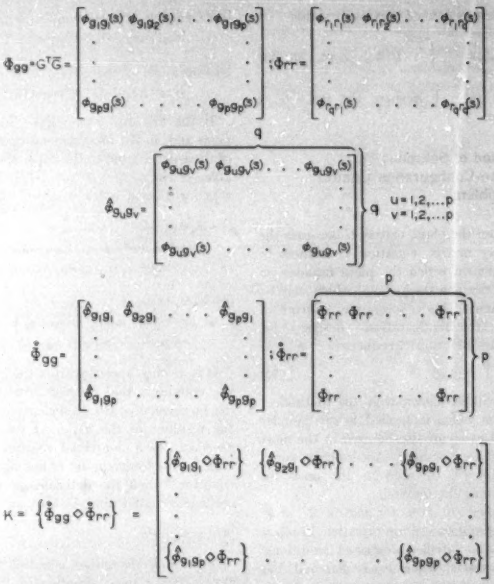


Fig. 3. Generation of effective coefficient matrix K for control problem

To preserve the form of equation 19, it is also necessary to construct new matrices for the right side of the equation. When this is done by a process similar to that just described, the reduction process is complete. Equation 11 thus may be rewritten as

$$[K\bar{A}]^{+} = [\bar{\Phi}_{rd}]^{+} \tag{23}$$

where K is the $pq \times pq$ coefficient matrix defined in the foregoing, and where $\bar{A}$ and $\bar{\Phi}_{rd}$ are the $pq \times 1$ matrices formed by arranging sequentially by rows the elements of A and $\bar{G}^{T}\Phi_{rd}{}^{T}$, respectively. In this form the matrix K plays the same role in the solution for all the elements of A as Φ_{rr} does in the filter or free-configuration solution for each row of A. Accordingly, the implicit method of solution can be applied directly to equation

23. It should be emphasized that it is not necessary to perform formally the reduction to equation 23 in order to obtain a solution to the multivariable control problem. For any specific problem, a direct expansion of equation 11 into its component equations automatically yields the coefficient matrix K.

Illustrative Example

To demonstrate the implicit method of solution one of the simplest 2-input–1-output free configuration examples will be examined. Since it has been shown that the most general control problem can always be reduced to this equivalent form, the method may equally well be applied to more complex configurations. The system is shown diagrammatically in Fig. 4. Each of the two inputs (r_j) is composed of a message (m_j) corrupted by white noise (n_j). The noises are uncorrelated with each other and with each message, while the message at the second input is specified to be the derivative of the message at the first input. It is desired to extract the best m_1 (in a mean-square sense) by operating on r_1 and r_2 simultaneously. The given data may be summarized as follows:

$$\phi_{r_1r_1}(s)=\phi_{m_1m_1}(s)+\phi_{n_1n_1}(s)$$
$$=\frac{3}{4-s^2}+1=\frac{7-s^2}{4-s^2}$$
$$\phi_{r_2r_2}(s)=\phi_{m_2m_2}(s)+\phi_{n_2n_2}(s)$$
$$=-\frac{3s^2}{4-s^2}+1=\frac{4-4s^2}{4-s^2}$$
$$\phi_{r_1r_2}(s)=\phi_{m_1m_2}(s)=\frac{3s}{4-s^2}$$
$$\phi_{r_jd_1}(s)=\phi_{m_jd_1}(s) \tag{24}$$

In the first step of the solution procedure the determinant of the matrix Φ_{rr} is set equal to zero as follows:

$$\begin{vmatrix} \frac{1}{(4-s^2)^2}\left|\begin{matrix}7-s^2 & +3s\\ -3s & 4-4s^2\end{matrix}\right| \end{vmatrix}=0 \tag{25}$$

Two "natural" system poles are provided, one at $s=-2$ and the other at

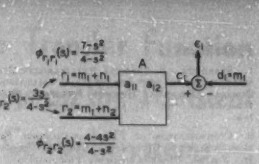


Fig. 4. Block diagram of illustrative example

$s=-1.32$. However, the pole at $s=-2$ is cancelled out by a corresponding zero thus leaving a single natural pole at $s=-1.32$. Now, since the left-half s-plane poles of the $\phi_{r_jd_1}(s)$ also appear in the $\phi_{r_jr_j}(s)$, no additional poles are present in the $a_{j1}(s)$. The form of the $a_{j1}(s)$ is thus

$$a_{11}(s)=\frac{p_{11}(s)}{s+1.32}$$
$$a_{21}(s)=\frac{p_{21}(s)}{s+1.32} \tag{26}$$

where the $p_{j1}(s)$ are undetermined polynomials in s.

A partial fraction expansion of the elements of the matrices of equation 11 may now be made. This indicates that the number of independent equations to be satisfied is two; therefore, it follows that $p_{11}(s)=x_{01}$ and $p_{12}(s)=x_{02}$ are simple gain constants. The two independent equations which result from equating residues are

$$5.25x_{01}-3.96x_{02}=0$$
$$3x_{01}-6x_{02}=2.04 \tag{27}$$

from which

$$x_{01}=0.411;\ x_{01}=0.545 \tag{28}$$

and

$$a_{11}(s)=\frac{0.411}{s+1.32};\ a_{12}(s)=\frac{0.545}{s+1.32} \tag{29}$$

Now, using equation 13 and substituting appropriately, the $\epsilon_1^2(t)_{\min}$ is computed to be 0.41.

A relative measure of this performance may be obtained by examining the best single-variable solutions for this example. If the best m_1 is extracted from r_1 alone, the use of equation 15 yields

$$a_{11}(s)=\frac{0.65}{s+2.63} \tag{30}$$

with a resulting

$$\overline{\epsilon_1^2(t)}_{\min}=0.65$$

Similarly, if the best m_1 is extracted from r_2 alone, the results are

$$a_{12}(s)=\frac{0.5}{s+1} \tag{31}$$

and

$$\overline{\epsilon_1^2(t)}_{\min}=0.50$$

Thus, the $\overline{\epsilon_1^2(t)}_{\min}$ of 0.41 for the multivariable solution compares favorably with the value of 0.50 which may be obtained by the best single-variable operation. As expected, the cost of this improvement is in the form of a more complex over-all transfer function.

References

1. General Algebraic Method Applied to Control Analysis of Complex Engine Types, A. S. Boksenbom, R. Hood. *Technical Report 980*, Lewis Flight Propulsion Laboratory, National Advisory Committee for Aeronautics, Cleveland, Ohio, 1950.

2. A Theory of Multidimensional Servo Systems, M. Golomb, E. Usdin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 253, no. 1, Jan. 1952, pp. 29–57.

3. The Synthesis of Multipole Control Systems, H. Freeman. *Technical Report T-15/B*, Electronics Research Laboratories, Columbia University, New York, N. Y., Apr. 1956.

4. The Application of Matrix Methods to Multi-Variable Control Systems, R. J. Kavanagh. *Journal*, Franklin Institute, vol. 262, no. 5, Nov. 1956, pp. 349–67.

5. The Theory of Multivariate Stochastic Processes, N. Wiener, P. Masani. *Acta Mathematica*, Stockholm, Sweden, vol. 98, 1957.

6. The Use of Canonical Expansions of Random Functions in Determining an Optimum Linear System, V. S. Pugachev. *Automation and Remote Control*, Moscow, USSR.

7. Design of Multivariable Optimum Filters, J. H. Westcott. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 80, no. 2, Feb. 1958, pp. 463–67.

8. The Linear Least Squares Synthesis of Continuous and Sampled Data Multivariable Systems, R. C. Amara. *Technical Report no. 40*, Stanford Electronics Laboratories, Stanford, Calif., July 1958.

Discussion

H. Freeman (Massachusetts Institute of Technology, Cambridge, Mass.): Dr. Amara extends Wiener's theory of optimum filter design to linear multivariable systems. The resultant matrix solution of the modified Wiener-Hopf equation represents a worth-while addition to our understanding of multivariable systems, and the author is to be congratulated on his efforts.

It is believed that a somewhat too restrictive point of view is taken when it is implied that the fixed plant matrix must be the identity matrix before the solution can reduce to that for a free configuration. Actually, the free-configuration solution will apply (at least in principle) for any plant matrix, provided the following three conditions are satisfied: 1. the inverse of the plant matrix exists, 2. the elements of the plant matrix have all their poles in the left-half s-plane, and 3. the zeros of the plant determinant lie in the left-half s-plane.[1]

The author states that he expects his technique to provide a practical synthesis procedure for multivariable control systems. The application of the optimization techniques based on the mean-square error criterion to control system synthesis has always been criticized for its supposed academic naïvete in ignoring the kernel of the control problem. The real objective of a control system is to obtain a specific, desired output behavior, precisely related to the inputs, in spite of the ever-present nonlinearities, time variations, and general

disturbances. For this reason also a practical control system without feedback is unthinkable, feedback being the necessary mechanism to remove the effects of whatever nonlinearities, time variations, or disturbances are inherently in the system. This is quite different from the problem of filter design, where feedback is not required per se. The author, instead of dealing with multivariable control systems, is actually describing what should be called a multivariable filter, and, indeed, his example also points this out.

This is not to imply that the application of the mean-square error criterion to a multivariable system is without merit; it is, however, only the first step to a possible practical synthesis procedure. The method would gain considerably in value if the optimization is made subject to the constraints that the mean-square values of the signals which are subject to saturation be held well within their linear range. This concept was introduced by Newton[3] for single-variable systems. Its extension to multivariable systems modifies equation 11 to the following form:

$$[\Phi_{rr}A^T(G^T\bar{G}+LG_s{}^T\bar{G}_s)]_+ = (\Phi_{rd}\bar{G})_+$$

where G_s = saturating signal transfer matrix, and L = diagonal Lagrangian multiplier matrix, and all other symbols are as used by the author.

Solution of this equation (i.e., finding the matrix A which satisfies it) will give a system transfer matrix which is optimum in the sense that it will have the minimum mean-square error consistent with the constraint that the probability of saturation be low. It is believed that some such procedure is absolutely necessary in the case of multivariable systems to give the optimization technique based on the mean-square error some practical significance.

It can also be pointed out that it is a relatively simple matter to extend the synthesis procedure described by the author to the case of deterministic signals (steps, ramps, etc.). This requires merely the use of the integral-square error rather than the mean-square error, with corresponding changes in the definitions of the power density spectra.

References

1. Stability and Physical Realizability Considerations in the Synthesis of Multipole Control Systems, **Herbert Freeman**. *AIEE Transactions*, vol. 77, pt. II, March 1958, pp. 1–5.

2. Compensation of Feedback-Control Systems Subject to Saturation, **George C. Newton, Jr.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 254, no. 4 and 5, Oct. and Nov. 1952, pp. 281–96, 391–413.

R. J. Kavanagh (University of Toronto, Toronto, Ont., Canada): The author approaches the problem of the synthesis of multivariable control systems incorporating feedback by designing an equivalent cascade system which may then be transformed into the corresponding feedback system. The process of transforming to the desired feedback system would appear to the writer to give rise to two difficulties in certain cases.

In the first place, while equation 3 is a valid relationship between A and A', there is no guarantee that A' is necessarily physically realizable. The criterion for realizability may be seen to be that the zeros of the determinant of $(I-GA)$ must be in the left half of the s-plane (disregarding any possible cancellations with right-half plane zeros of A).

The second difficulty arises in connection with the free configuration control problem. This problem occurs when the plant transfer matrix G becomes the identity matrix. If G is the identity matrix, this implies that either $p=q$ or that G is a $p\times p$ matrix. If $p=q$, there is no problem other than that of realizability since the cascade system is readily converted to a feedback system. However, if G is a $p\times p$ matrix and $p\neq q$, the problem is not solvable in terms of a feedback system since the number of outputs of G and the number of inputs to A' (and to A) must be equal. Hence it is the writer's belief that the procedure for the free-configuration case, while valid for any cascade system, is only valid for feedback systems when $p=q$.

R. C. Amara (Stanford Research Institute, Menlo Park, Calif.): Dr. Freeman is entirely correct in pointing out that the free-configuration solution may be applied when the three conditions he enumerates for the plant matrix are satisfied. The author did not intend to imply that the form of equation 14 results only when G is the identity matrix. The particular cases in which equation 11 is reducible to the form of equation 14 are discussed elsewhere (see reference 8 of the paper).

The extension of the synthesis procedure to optimization subject to constraints on the mean-square values of plant input saturation signals is indeed commendable. The possibility and desirability of including such a constraint in the procedure for the multivariable case has been previously suggested in reference 8 of the paper. However, the statement that the synthesis procedure described deals only with multivariable filters is not believed to be correct. Feedback is included as an integral part of the control systems considered. The particular multivariable filter example which is used as an illustration was chosen primarily for its simplicity. The application of the method to the solution of multivariable control examples is included in reference 8 of the paper.

It may also be noted that, in addition to the extension of the procedure to deterministic signals as indicated by Dr. Freeman, it is also a relatively simple matter to adapt the method to synthesis for minimization of the sum of the weighted-output mean-square or integral-square errors.

Dr. Kavanagh properly indicates that the solution of equation 3 does not guarantee the physical realizability of matrix A'. When the zeros of the determinant $(I-GA)$ do not all lie in the left half of the s-plane, a $q\times q$ matrix B may be inserted in the feedback loop. This will introduce sufficient freedom to satisfy the condition that all the zeros of the determinant $(I-BGA)$ be in the left-half s-plane. A similar procedure is used in single-variable systems.

Finally, as indicated previously, the free-configuration case as represented by equation 14 may be applied whenever the three conditions stated by Dr. Freeman are satisfied. Since G^{-1} must exist, then p must equal q for this case. When p is not equal to q, the problem is solvable as in the general control case.

Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems

OLLE I. ELGERD
MEMBER AIEE

WILLIAM C. STEPHENS
NONMEMBER AIEE

THE ROOT LOCUS methods have in recent years gained a widespread popularity among control engineers engaged in the work of both analysis and synthesis. The reason for this is rather obvious; the root locus techniques give the system designer more information of the kind in which he is interested when analyzing a given system; when he is synthesizing a control network, these techniques permit him to build the desired characteristics into the system in a much more direct fashion than did the old frequency response techniques. The basis for this statement is that in the large majority of control systems, the transient, not the frequency, response is of overshadowing importance.

To understand better the aim of this paper, it will be necessary to review briefly the basic idea of the root locus methods as described in the literature[1-4] (The reference list is by no means complete but contains a representative selection of the most popular techniques.) For the purpose of demonstration, a simple 1-loop system having a forward loop transference $G(s)$ and a feedback loop transference $H(s)$ is chosen for the discussion.

The transient response of this system to an input signal $r(t)$ can be written

$$c(t) = \mathcal{L}^{-1}\left(R(s)\,\frac{G(s)}{1+G(s)H(s)}\right) \qquad (1)$$

or

$$c(t) = \mathcal{L}^{-1}[R(s)\times K(s)] \qquad (2)$$

In the second expression is introduced the closed-loop transference

$$K(s) = \frac{G(s)}{1+G(s)H(s)}$$

If attention is limited to true follow-up systems of lumped type, i.e., systems with zero position error for a step input, then it is possible, as is easily verified by means of the "final value theorem," to put $K(s)$ in the following form:

$$K(s) = \frac{\prod_{i=1}^{n}(-P_i)\;\prod_{\mu=1}^{m}(s-Z_\mu)}{\prod_{i=1}^{m}(-Z_i)\;\prod_{\nu=1}^{n}(s-P_\nu)} \qquad (3)$$

The Z_μ's are the closed-loop (and at the same time open-loop) zeros and P_ν's the closed-loop (but not the open-loop) poles.

By choosing as a test signal a unit step input, equation 2 will give after application of Heaviside's expansion theorem

$$c(t) = 1 + \sum_{\nu=1}^{n} A_\nu e^{P_\nu t} \qquad (4)$$

The amplitudes A_ν can be calculated from

$$A_\nu = \frac{1}{P_\nu}\,\frac{\prod_{i=1}^{n}(-P_i)}{\prod_{i=1}^{m}(-Z_i)}\,\frac{\prod_{i=1}^{m}(P_\nu-Z_i)}{\prod_{i=1}^{\nu-1}(P_\nu-P_i)\;\prod_{i=\nu+1}^{n}(P_\nu-P_i)} \qquad (5)$$

The closed-loop poles and zeros may be either real or conjugate complex but for any stable, physically realizable system they will contain negative real parts. It is also clear from equation 4 that all terms appearing under the summation symbol represent the transient part of the time response. These terms are either monotonously decaying (in the case where the corresponding poles are real) or oscillatorily decaying (should any poles appear as conjugate complex). Note that the zeros will affect only the magnitude of the amplitudes.

Equations 4 and 5 reveal one very important fact: The time response (once the input is known) is uniquely determined by the location of the closed-loop poles and zeros in the s-plane. All root locus techniques are primarily based on this very fact. The designer manipulates the circuit parameters either in the forward or feedback path in such a manner that the closed-loop zeros and poles will form an s-plane configuration corresponding to the particular response desired. Unfortunately, this is easier to say than do.

There are three main hurdles that the designer must overcome:

1. He must find the numerical values of the poles and zeros of the open-loop transfer functions. Often he may be able to work under the assumption of noninteraction between the different blocks of the system. In such a case the problem of finding these poles and zeros is quite simple.

2. Using semigraphical techniques, he will investigate how the closed-loop poles and zeros travel throughout the s-plane when system parameters are subject to changes. This part of the over-all problem has been lucidly treated in the literature. Some more or less complex devices, for example, the Spirule, have been designed to help out with this task.

3. The most difficult problem is the translation of the pole-zero configuration into time response. Principally, this is an easy task. After having made a parameter adjustment and after having located the new positions of the closed-loop poles and zeros, one simply utilizes the two equations, 4 and 5. One should remember, however, that to get full information in the real time domain it is necessary to calculate the output for a sequence of time values. It is also worth remembering that if the result is not satisfactory the same procedure has to be repeated once or several times depending upon how fast the result converges toward an acceptable response.

Obviously, a digital computer is a very helpful aid, but from a financial point of view it is seldom feasible to employ such a device. To perform the necessary computations with a slide rule or with a desk calculator is in most cases extremely time consuming. This particular disadvantage of the root locus methods has quite possibly caused many a control system engineer to take a reserved attitude toward these techniques.

Objectives

The material presented in this paper is intended to be of help for the designer who utilizes the root locus techniques in his work. The bulk of the information is given in the form of graphs which are intended to serve as a table for the designer. By means of these graphs he will be able to correlate immediately the connection between pole-zero configura-

Paper 59-107, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 27, 1958; made available for printing December 3, 1958.

OLLE I. ELGERD is with the University of Florida, Gainesville, Fla., and WILLIAM C. STEPHENS is with the U. S. Army Combat Surveillance Agency, Arlington, Va.

The paper reports the results from an investigation performed at the University of Florida Electrical Department.

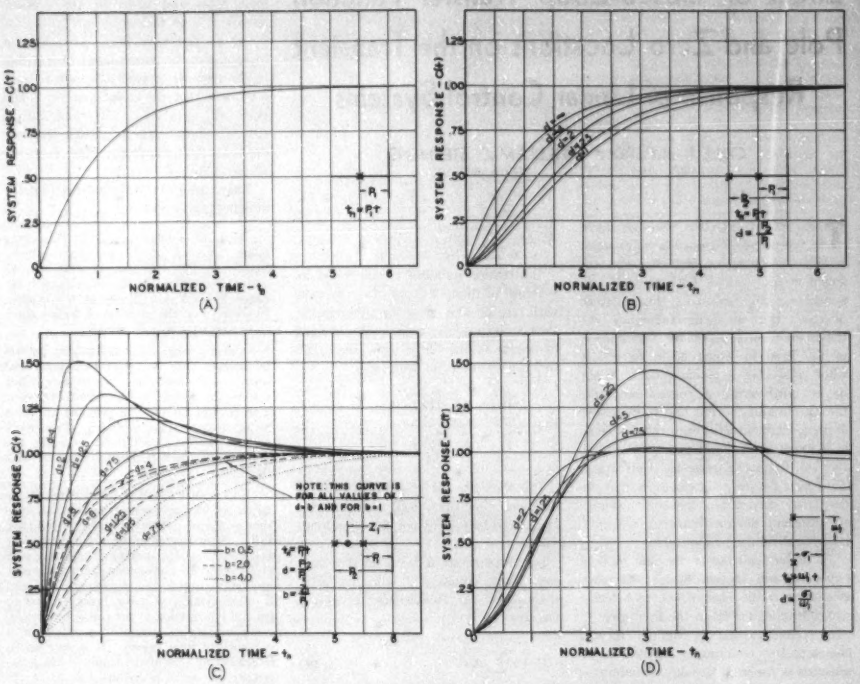


Fig. 1. Responses to a step input of systems of varying characteristics

A—Having one negative real closed-loop pole only
B—Having two negative real closed-loop poles
C—Having two closed-loop poles and one zero; all real and negative
D—Having two conjugate complex poles

tion and real-time response. They serve, in other words, in a similar way, as does a table of Laplace transforms; from such a table one finds the real-time response associated with a particular function in the complex *s*-domain.

A Laplace transform table will always be characterized by a certain degree of incompleteness. To make it complete it would be necessary to include an infinite number of transform pairs. It is appreciated however, that even a limited Laplace transform table is extremely helpful because it will contain the types of functions most frequently encountered.

The same general statements apply to the graphs in this paper. It would seem at first glance to be a prodigious task to try to investigate and tabulate every possible pole-zero configuration, of which there exists an infinite number. How this difficulty has been circumvented will be shown later; first, the earlier work done in this general area will be briefly surveyed.

Review of Previously Published Work

No attempt has been made to include here the many papers concerned with obtaining the transient response from a knowledge of the frequency response because this is, as already stated, not the objective of this paper. Only those treatises will be mentioned where a direct relationship is established between system transfer function pole and zero locations and time response. Not very much information is available in the technical literature.

One of the most fundamental pole-zero configurations is the one consisting of two conjugate complex poles. It is natural that this case first appeared in the literature. In 1947, James, Nichols, and Phillips[5] presented a series of graphs showing the unit-step-function response for such a system. Truxal,[6] in 1955, added a negative real zero to the conjugate complex pole pair and calculated some data for this configuration.

Mulligan had already in 1948 in a doctorate dissertation[7] made an extensive study of the connection between the

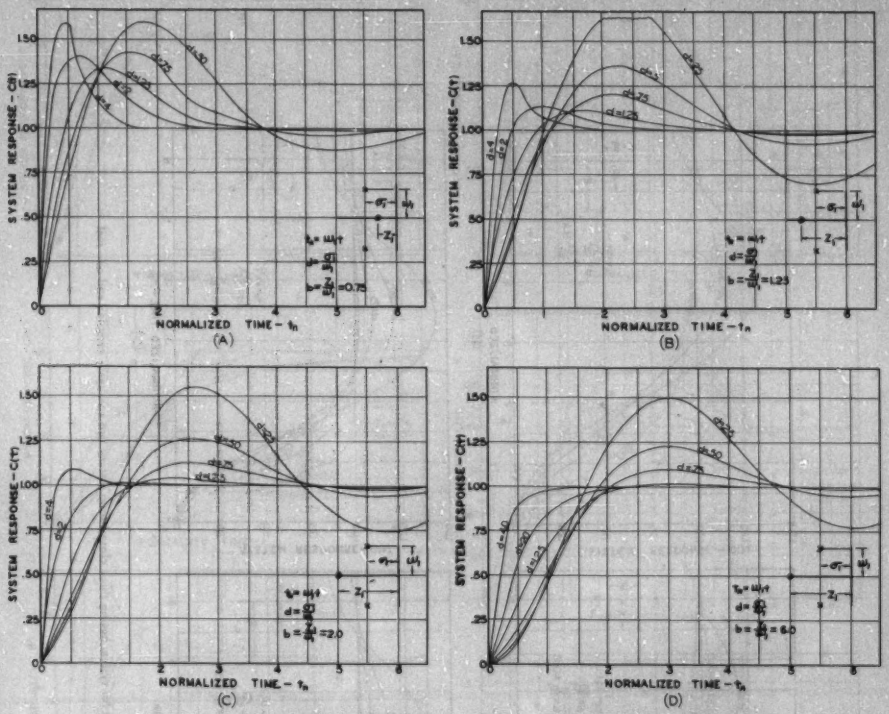


Fig. 2. Responses to step input for system having two closed-loop conjugate complex poles and one negative real zero

pole and zero locations and the maxima and minima of the transient response. He was particularly interested in the response of multistage, nonfeedback amplifiers, but because of his general approach, his results have had some application also in the control field.

Zemanian in 1955 in his doctorate dissertation[8] extended the work of Mulligan to include a method of finding that value of time at which the response crosses the final value line. Elgerd introduced on a very limited scale the ideas of the present paper into the field of sampled-data systems in a paper[9] appearing in July 1958. An attempt was made to classify sampled-data systems on the basis of the closed-loop pole and zero configurations.

A recent paper by Huang[10] takes up for discussion the sensitivity of the system response of the pole locations. In particular, he co-ordinates the constancy of response of a closed-loop system by a definition of the sensitivity of a closed-loop pole as the percentage variation of this pole for a given percentage variation of one of the open-loop parameters.

Fundamental Pole-Zero Configurations

One of the main problems in this investigation has been to extract from the infinite number of pole-zero configurations the ones that have a practical significance. When, after a careful sampling procedure, these fundamental cases have been selected, there remains the problem of bringing to a minimum the number of independent graphs needed to cover each case adequately. This task of system classification is greatly simplified due to a fundamental characteristic feature of linear systems, namely, the predominance of certain poles and zeros. A study of equation 5 will reveal that those amplitudes A_r that are associated with poles P_r, located comparatively far away from the origin in the s-plane, will be of relatively small magnitude. Not only do these transient terms in equation 4 have small magnitudes but they will also die out very fast. One might, therefore, without losing much accuracy, obtain the response of a specific system by considering only the poles and zeros located closest to the origin. It is difficult to give a general rule of thumb as to where to put the limit. Truxal[11] mentions a criterion sometimes used in design.

The following configurations have been

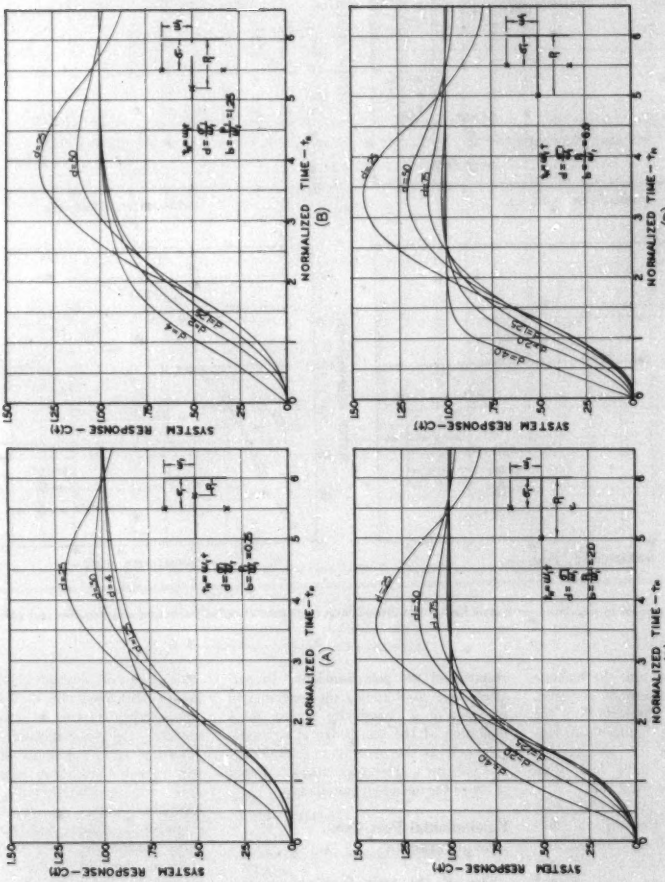


Fig. 3. Responses to step input for system having three closed-loop poles: two conjugate complex and one real

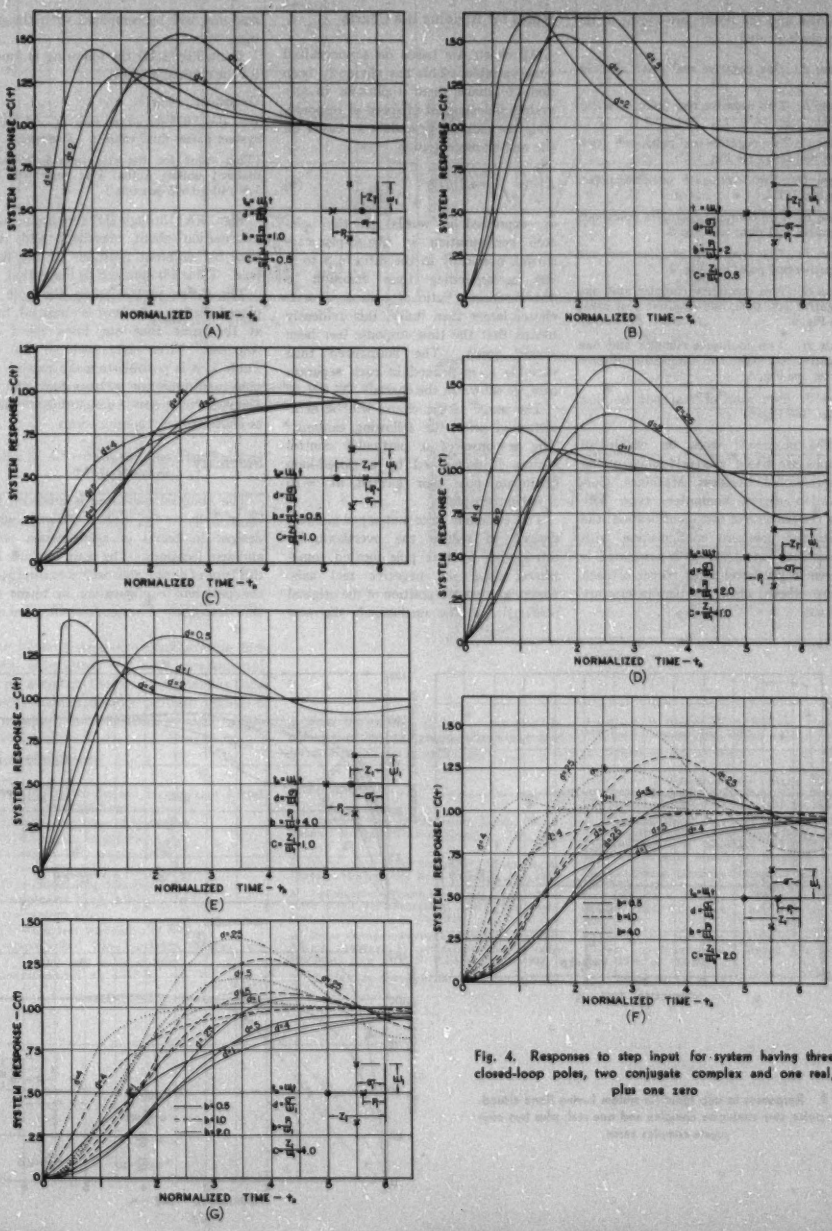


Fig. 4. Responses to step input for system having three closed-loop poles, two conjugate complex and one real, plus one zero

studied and the results are shown in the graphs indicated:

Case *A:* One negative real pole; see Fig. 1(A).

Case *B:* Two negative real poles; see Fig. 1(B).

Case *C:* Two negative real poles, one negative real zero; see Fig. 1(C).

Case *D:* Two conjugate complex poles; see Fig. 1(D).

Case *E:* Two conjugate complex poles, one negative real zero; see Fig. 2.

Case *F:* Two conjugate complex and one negative real pole; see Fig. 3.

Case *G:* Two conjugate complex and one negative real pole, one negative real zero; see Fig. 4.

Case *H:* Two conjugate complex and one negative real pole, two conjugate complex zeros; see Fig. 5.

Case *I:* Two pairs of conjugate complex poles; see Fig. 6.

The numerical work on which all graphs are based was performed on an International Business Machines Corporation digital computer, type *650*. For the purpose of fast identification, the particular pole-zero configuration with which each set of curves is associated is shown in the lower right corner of each figure where, also, all parameters are defined.

Rules for Reading the Charts

All charts are based on a normalized time variable. This has obviously been done to make them applicable to any system independent of speed of response. This normalizing procedure is based on the well-known equation

$$\mathcal{L}\left[f\left(\frac{t}{a}\right)\right] = a\mathcal{L}[f(t)] \qquad (6)$$

or, expressed in words: If the pole-zero configuration in the s-plane is shrunk uniformly in the ratio of a to 1, the corresponding time function is stretched in the ratio of 1 to a. If a is chosen larger than unity, this evidently means that the time response has been slowed down. The normalized time variable t_n is defined in each separate case, as shown in the charts.

The usage of the charts will be demonstrated with the following example: The response of a particular control system is dominated by a closed-loop conjugate pole pair located at $-10 \pm j40$ in the s-plane.

This system is quite undamped and it is desired to reduce the overshoot by introducing another pole located somewhere along the negative real axis. Assuming that the position of the original pole pair will be unchanged, the new response will be compared with the old one.

From Fig. 1(D) the following is found for the original system:

Overshoot: 45%

Rise time (defined as the time at which the system passes final value): 1.8 seconds

(This value for the rise time is in normalized units. Actual rise time will be 1.80/40 = 0.045 seconds.)

Figs. 3(A) through (D) give immediate information about overshoot and rise time for different positions of the real pole. Table I is obtained by inspection.

The closer to the origin the pole is placed the less overshoot is obtained, but at the same time one loses speed of response. From data such as that of Table I, it is possible to make an acceptable compromise or perhaps find that a different pole-zero configuration must be chosen to meet specifications.

Summary

The material presented is intended to be of help to the system designer who designs in terms of transference pole and zero locations. The main obstacle in this type of design approach is to interpret the pole-zero configuration in terms of time response. A system classification on

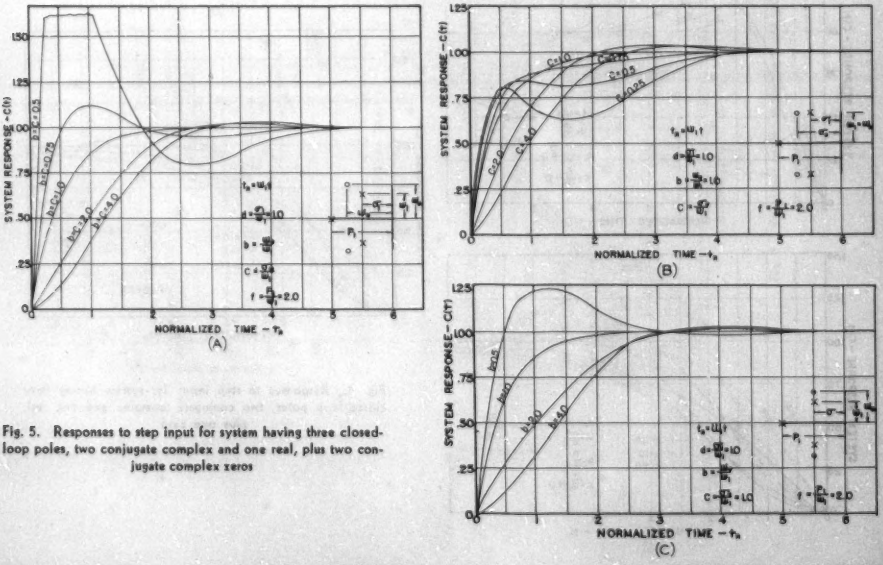
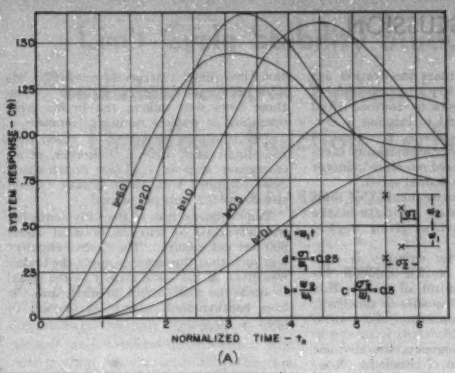
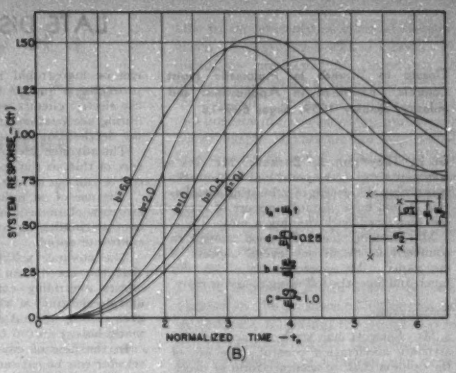


Fig. 5. Responses to step input for system having three closed-loop poles, two conjugate complex and one real, plus two conjugate complex zeros





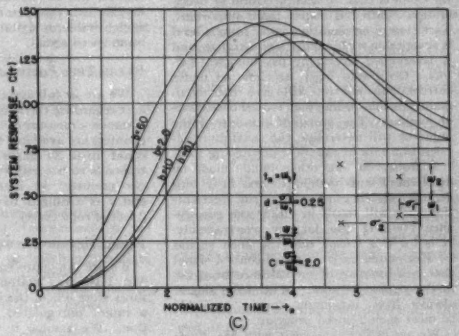


Fig. 6. Responses to step input for system having one pair of conjugated complex closed-loop poles and one pair of zeros

Table I. Overshoot and Rise Time for Real Pole Variation

Pole Location	Overshoot, Per Cent	Rise Time Seconds
−30	19	3.2
−50	32	2.65
−80	39	2.35
−240	44	2.0

the basis of pole-zero configuration has been attempted; for each fundamental type of system the time response has then been calculated and plotted versus a normalized time variable.

Nomenclature

A_r = amplitudes defined in equation 4 and calculable from equation 5
$c(t)$ = time response of system
$C(s)$ = Laplace transform of $c(t)$
$G(s)$ = open-loop transference
$H(s)$ = feedback loop transference
$K(s)$ = closed-loop transference
P_ν = poles of $K(s)$; $\nu = 1,2,\ldots n$
$r(t)$ = system input
$R(s)$ = Laplace transform of $r(t)$
t_n = normalized time variable, defined for each case in the graphs
Z_μ = zeros of $K(s)$; $\mu = 1,2,\ldots m$

Some parameters, a, b, c,... are used in connection with the graphs where they also have been defined in each case.

References

1. Control System Dynamics (book), W. R. Evans. McGraw-Hill Book Company, Inc., New York, N. Y.

2. Control System Synthesis (book), L. G. Truxal. McGraw-Hill Book Company, Inc., 1955.

3. Integrated s-Plane Synthesis Using 2-Way Root Locus, John Zaborszky. *AIEE Transactions*, (Jan. 1957 section), pp. 797–801.

4. Root Locus in Feedback System Synthesis, John A. Aseltine. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Mar. 1956, p. 402.

5. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, Inc., 1947, p. 143.

6. Truxal, op. cit., pp. 40–41.

7. The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems, L. H. Mulligan, Jr. *Proceedings*, Institute of Radio Engineers, vol. 37, May 1949, pp. 516–29.

8. Further Effects of the Pole and Zero Locations on the Step Response of Fixed Linear Systems, Armen H. Zemanian. *AIEE Transactions*, vol. 74, pt. II, March 1955, pp. 52–55.

9. An Analog Computer Study of the Transient Behaviour and Stability Characteristics of Serial-Type Digital Data Systems, Olle I. Elgerd. *Ibid.*, vol. 77, pt. I, July, 1958, pp. 358–366.

10. The Sensitivity of the Poles of Linear Closed-Loop Systems, R. Y. Huang. *Ibid.*, pt. II, Sept. 1958, pp. 182–87.

11. Truxal, op. cit., p. 43.

LATE DISCUSSION

Closure to "Trends in Automatic Hoist Controls," published in Applications and Industry, January 1959, pages 609–18.

G. L. Tiley and **E. Zucker**: Mr. Ivy's comments are to the point and his remarks concerning the design of a hoisting installation should have the full support of everyone concerned with running a mine.

Mr. Eastcott has been good enough to comment in detail on several aspects of our paper. We appreciate his effort very much and have the following to say in reply.

Regulators

Mr. Eastcott has stressed the need for accurate acceleration control. We are of the opinion that an accurate form of deceleration control is much more important. Variation of acceleration away from a level is acceptable within quite a wide range, and is determined largely by psychological factors. On a skip hoist, acceleration under current-limit control will give the maximum possible production for all conditions of loading. The speed of run-out of the inductor will determine the maximum acceleration rate in case of a downgoing load.

A hoist, on the other hand, must be decelerated and brought to rest in a very definite distance. Variation in deceleration rate will result in either the conveyance creeping too far, or overtraveling. It is for this reason that the hoist should be decelerated under a speed-control signal that is a function of the distance from the final stopping position. The drive should always have maximum torque available to meet such a speed program.

We are in full agreement with Mr. Eastcott on the desirability of using a rate feedback for stabilizing, especially on a current-limit control where a double-rate feedback is most effective.

The use of a tachogenerator voltage or a generator voltage feedback is largely a matter of convenience. Both can be made to give accurate speed control. In the former case, some complications arise in suiciding at standstill. In the latter case, different values of current-resistance compensation have to be used for creep and full speed.

Program Devices

The use of simple limit switches, in place of the advance selector, does not necessarily result in simplification of the control circuits. If maximum production is required, even on a single-level single-skip hoist, two sets of limit switches are needed (one for hoisting and one for lowering) because: 1. acceleration away from a level can be more rapid than deceleration approaching it; and 2. the time lags, etc., of the electric circuits work to decrease speed during acceleration and to increase it during deceleration.

The advance selector automatically takes care of this, as the speed reference inductor can be run out at any desired speed.

The use of an inductor instead of limit switches eliminates many contacts in the control circuits, as well as giving stepless control of speed.

The advance selector circuitry is such that the speed cam switch inductor exercises overriding control of acceleration and deceleration at all points in the shaft, even during manual operation of an automatic hoist.

In the case of emergency, the advance selector can be cut out of circuit by reconnecting two wires, to replace the speed cam switch inductor signal by a constant maximum speed signal.

Brake Test Equipment

We are in full agreement with Mr. Eastcott regarding the relative usefulness of the distance - measuring device. Mechanical counters are available that will satisfactorily count up to 25 pulses a second. This corresponds to one pulse per foot at 1,500 feet per minute. At higher speeds, the circuitry is arranged to count one pulse per 2 feet of rope travel.

Brake Design and Performance

We are not quite sure what is meant by Mr. Eastcott's statement that the brake shoes store 4% of the impact energy during a rapid, unregulated, emergency application. We assume it to mean that, on a weight-applied brake, the dead weight stores 96% of the impact energy. Under these circumstances, the elimination of the dead weights would allow the application time to be speeded up by a factor of $\sqrt{24}$ or 4.9.

About the minimum application time that can be achieved without excessive overshoot by a weight-applied brake is 2 seconds. Therefore we would not expect that an air-applied caliper-type brake, adjusted to normal operating clearances, can be applied in better than 0.4 second without excessive overshoot. If an application time of 0.2 second is required, the inertia would have to be reduced to 1%.

We have conducted many tests on brake disks under all conditions of operation and are convinced that disks will prove far more satisfactory brake paths than the old ones from both the mechanical and thermal points of view.

We have witnessed acceptance tests on six hoists equipped with air-applied Microspot caliper brakes, and the brake application time under emergency conditions has been between 0.8 and 1.2 seconds. When these tests were taken, the brakes were operating at normal running clearances.

Fig. 1 of this closure is a speed–time recording, taken with a Visicorder, of an emergency stop on a 14-foot double-drum hoist equipped with these brakes. Chart speed was 0.2 inch per second.

The overspeed cam of the Lilly controller was advanced to trip the hoist at about 800 feet per minute. The control circuitry was such that the signal to apply the brakes was given at the same instant as the signal to open the loop circuit breaker, and the loop breaker took about 0.1 second to open.

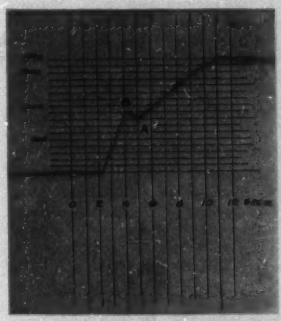


Fig. 1. Visicorder speed–time recording of air-operated caliper brakes on 14-foot double-drum hoist

At *A*, the hoist started to speed up, signifying that the loop breaker had opened. Speed increased due to the out-of-balance rope torque until *B*, where the brakes started to apply. Dead time was about 0.8 second, i.e., the actual brake application time was about 0.9 second. The difference between the times we observed and the times recorded by Mr. Eastcott, probably can be accounted for by the clearance necessary in practice between the shoes and the brake drum. This clearance prevents the brakes binding when the drum heats up, and mine operating personnel keep it at safe values.

In conclusion, we are in full agreement with Mr. Eastcott on the desirability of priming to reduce impact shock loads on the brake mechanism.

Limit-Cycle Stability Study of a Feedback Control System by a New Describing-Function Technique

H. J. HARRINGTON
ASSOCIATE MEMBER AIEE

A NEW describing-function technique is used in this paper to show the effects of several gain-controlling nonlinearities on the limit-cycle stability of an aircraft roll-stabilization system. The new method leads to a stability diagram which is divided not into the usual two regions of "stable" and "unstable" operation but rather into numerous regions each of which is associated with a particular type of limit-cycle oscillation or singular point. Such a diagram gives considerable insight into the mechanism of oscillations often observed in physical equipment. Very satisfactory agreement with analog computer studies is demonstrated for the new technique.

The case of simultaneous limiting in each of two feedback loops is easily handled for the system studied herein. A logical synthesis procedure for deliberately introducing a new nonlinearity in order to counteract an existing one is also proposed.

A paper[1] treating the stability of linear systems furnished the impetus for this work; another paper[2] concerned itself with some of the same problems which are treated here but the approach and interpretation of results were considerably different.

Although the technique developed here may be employed for any nonlinearity for which a describing function can be derived, it appears to offer interpretative advantages over the more conventional approach[3] only in the case of gain-controlling nonlinearities; that is, nonlinearities which do not introduce phase shift to signals passing through them.

Paper 59-233, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted September 29, 1958; made available for printing December 22, 1958.

H. J. HARRINGTON was formerly with Convair-Astronautics, A Division of General Dynamics Corporation, San Diego, Calif., and is now with Solar Aircraft Company, San Diego, Calif.

The author wishes to thank Mr. Carl J. Scherer of McDonnell Aircraft Corporation for first using the methods of this paper in a practical application and also for obtaining the analog computer checks presented herein.

Nomenclature

A = peak amplitude of a sinusoidal oscillation in roll rate
d = value of threshold in rate gyroscope
$E_m = M_d - M_t$, error in normalized rolling moment
f = viscous damping coefficient of air frame
I = moment of inertia of air frame about roll axis
i = control surface incidence
i_c = control surface incidence command
$j = \sqrt{-1}$
$K_a = M_t/i$, aerodynamic effectiveness of control surface
K_p = position feedback gain
K_r = rate feedback gain
L_p = position-gyro limit
L_r = rate-gyro limit
M_d = rolling moment (normalized to I) caused by external disturbance
M_t = rolling moment (normalized to I) caused by control surface incidence
N_p = equivalent gain, or describing function, of nonlinearity in position gyro
N_r = equivalent gain, or describing function, of nonlinearity in rate gyro
s = Laplace operator
T_1, T_2 = constants of control servo
$\alpha_d = d/A$, amplitude parameter for threshold describing function
$\alpha_l = L_r/A$, amplitude parameter for limiting describing function
$\gamma = L_p/L_r$, position-to-rate limit ratio
ϕ = roll-position angle
$\Omega = \gamma\omega$, normalized frequency
ω = circular frequency

Description of System Studied

The system to be studied, which is shown in block diagram in Fig. 1, is a typical aircraft roll-stabilization system used to maintain air-frame roll angles near zero in the presence of external disturbances. The system works as follows. A disturbing moment M_d produces a moment error which in turn causes the air-frame to roll. The instantaneous roll rate and position are measured by gyroscopes whose output signals are suitably scaled by the gains K_r and K_p, added, and sent to a control surface servo as an incidence command i_c. The output incidence of the control servo then gives rise to a control moment M_t, which tends to reduce the moment error and thus stabilize the air-frame.

This seemingly simple system is complicated by the presence of limiting in, or immediately following, the position gyro and by both limiting and threshold in the rate gyroscope. In addition, the aerodynamic gain K_a is subject to rather wide variations from nominal because of unpredictable changes in aircraft flight conditions; good stability margins with respect to the gain K_a must therefore be attained to prevent either divergent oscillations or a hard limit cycle of catastrophic proportions.

In the study which follows, it will be assumed that the linearized transfer function of both gyroscopes is unity. Although this might well be a good assumption, the restriction is not at all necessary and is made only for the sake of brevity as is the omission of compensating networks in either of the two feedback paths. The nominal system parameters for the rest of the system are as follows:

$$\left.\begin{array}{l} f/I = 0.67 \text{ per second} \\ T_1 = 0.04 \text{ second} \\ T_2 = 0.05 \text{ second} \end{array}\right\} \quad \left.\begin{array}{lll} K_a & d & L_p \\ K_p & & L_r \\ K_r & & \end{array}\right\} \text{ general} \qquad (1)$$

Generalized Stability Chart

The characteristic equation $F(s)$ of the system of Fig. 1 is

$$F(s) = T_1T_2s^4 + \left(T_2 + T_1\frac{f}{I}\right)s^3 + \left(1 + T_2\frac{f}{I}\right)s^2 + \left(K_aK_rN_r + \frac{f}{I}\right)s + K_aK_pN_p \qquad (2)$$

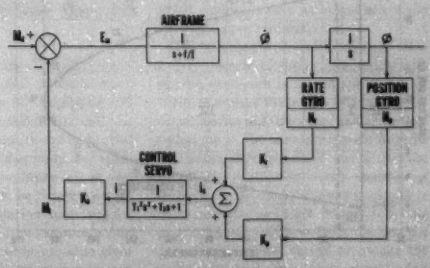


Fig. 1. Block diagram of an aircraft roll-control system

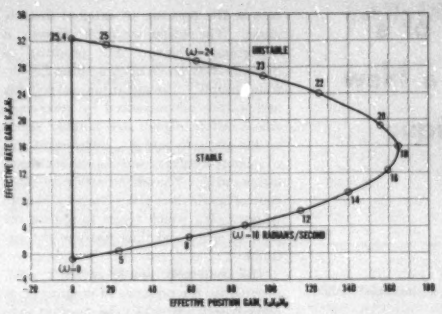


Fig. 2. Stability diagram for the roll-control system

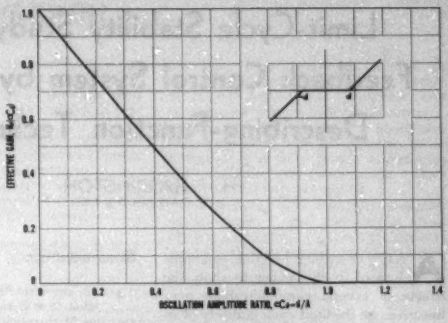


Fig. 3. Describing function for threshold

For the present the nonlinearity terms N_p and N_r are assumed to be simple constants. Setting equation 2 equal to zero and replacing s by $j\omega$ gives an equation which represents the marginally stable case if real roots in ω exist:

$$T_1^2\omega^4-\left(1+T_2\frac{f}{I}\right)\omega^2+K_aK_pN_p- j\omega\left[\left(T_2+T_1^2\frac{f}{I}\right)\omega^2-\left(K_aK_rN_r+\frac{f}{I}\right)\right]=0 \quad (3)$$

But for equation 3 to be valid, both the real and imaginary parts must be identically zero; therefore, this equation may be replaced by two separate equations from which the system gains can be solved

$$K_aK_pN_p=\left(-T_1^2\omega^2+1+T_2\frac{f}{I}\right)\omega^2 \quad \text{(real part)} \quad (4)$$

$$K_aK_rN_r=\left(T_2+T_1^2\frac{f}{I}\right)\omega^2-\frac{f}{I} \quad \text{(imaginary part)} \quad (5)$$

For the particular parameter values in equation 1, these equations become

$$K_aK_pN_p=(-0.0016\omega^2+1.033)\omega^2 \quad (6)$$

$$K_aK_rN_r=0.0511\omega^2-0.67 \quad (7)$$

A plot of equations 6 and 7 with ω treated as a real-value parameter is shown in Fig. 2. By applying Hermite's criteria[4] (or any other convenient method), it can be shown that the region within the closed boundary is stable while that outside the boundary line is unstable. This situation is approximately true even when the system is nonlinear if the effective gains of the nonlinearities N_p and N_r are explicitly included in the gain coordinates as shown in Fig. 2.

Limit-Cycle Stability Chart

For better interpretation of the nonlinear cases, it is desirable to have the stability plot in terms of just the linear gains K_aK_p and K_aK_r, and show not only the wholly stable and unstable regions but also the various limit-cycle oscillation conditions which might arise. To illustrate this, the case of a single nonlinearity, namely, threshold in the rate gyro, will be examined.

It will be assumed that the effective gain of the threshold element N_r is given by its describing function[2] during an oscillation. A plot of N_r as a function of normalized roll-rate amplitude α_d is shown in Fig. 3. Note that if the roll-rate amplitude is equal to or less than the deadspot, that is $\alpha_d \geq 1$, the gain N_r is zero. As the roll-rate amplitude approaches infinity, however, the grain N_r approaches unity and the system should behave linearly.

The consequences of this nonlinearity may be seen by examining Fig. 4. Here the solid curve is identical to that of Fig. 2 while the broken lines, which reflect the presence of the threshold, divide the over-all chart into regions of varying degrees of stability. In general, linear-gain co-ordinates appearing as a point in Fig. 4 will be transformed into a vertical line in Fig. 2. Since N_p is by definition unity, this line will be located at the same abscissa and will descend from the point ordinate ($N_r=1$) to the position-gain axis

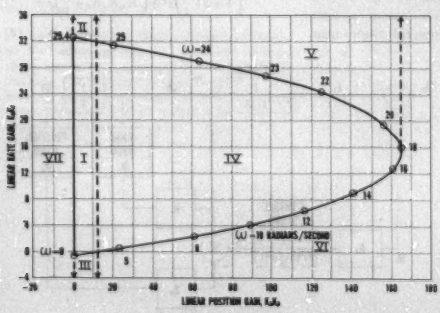


Fig. 4 (left). Limit-cycle stability diagram for case of threshold in rate gyroscope

REGION	FREQ. RAD./SEC.	STABILITY DESCRIPTION	PHASE-PLANE SKETCH
I	—	STABLE SINGULAR POINT	
II	> 25.2 < 25.4	HARD, UNSTABLE LIMIT CYCLE	
III	> 0 < 3.0		
IV	> 3.0 < 10	SOFT, STABLE LIMIT CYCLE	
V	> 3.0 < 10	SOFT, STABLE LIMIT CYCLE	
	> 10 < 25.2	HARD, UNSTABLE LIMIT CYCLE	
VI & VII	—	UNSTABLE SINGULAR POINT	

Fig. 5 (right). Phase-plane sketches showing typical trajectories associated with the seven regions of Fig. 4

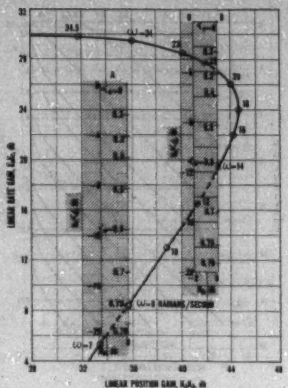


Fig. 6 (left). Logarithmic plot of limit-cycle stability diagram showing overlay of describing function for rate-gyro threshold

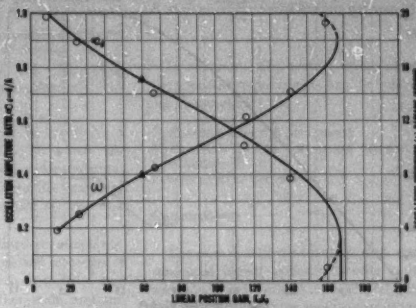


Fig. 7. Amplitude and frequency of limit-cycle oscillations caused by rate-gyro threshold for variations in position gain K_aK_p, while holding rate gain constant at $K_aK_r = 20$

—— Stable oscillations as determined by describing functions
– – – Unstable oscillations as determined by describing functions
○ Points obtained by analog computer

($N_r = 0$). If the linear-gain co-ordinates fall in region I of Fig. 4, no amount of attenuation because of $N_r(\alpha_d)$ can cause the corresponding gain ordinate in Fig. 2 to fall in the unstable region. Thus, for region I, any initial disturbance must ultimately die out as is depicted by the corresponding phase-plane sketch of Fig. 5. (The phase-plane sketches of Fig. 5 are for purposes of qualitative description only since a polydimensional phase space is required to accurately describe higher-order systems.)

If the linear-gain co-ordinates fall in either regions II or III, it is possible to excite an unstable limit-cycle oscillation. This behavior may be predicted by noting from Fig. 2 that the system is stable when at rest ($N_r = 0$), and unstable when excited sufficiently to cause the gain $K_aK_rN_r$ to exceed the stable-region ordinate.

Region IV of Fig. 4, which is the most probable region of operation for the system in question, leads to a soft, stable limit cycle. For examination of Figs. 2 and 3 shows that with the system at rest the effective rate gain is zero and the system is unstable; but, as the amplitude of the resulting divergent oscillation builds up, α_d decreases and $N_r(\alpha_d)$ increases until the gain reaches the lower stability boundary of Fig. 2. Further increases in the oscillation amplitude force the effective gain ordinate into the stable region with the result that the excess amplitude is damped out. Obviously the frequency of oscillation must lie in the range of $3.6 < \omega < 18$ radians per second.

Both a soft, stable limit cycle and a higher-amplitude unstable oscillation are possible if the system gains are adjusted to lie in region V of Fig. 4. Repetition of the preceding logic will show that the stable oscillation will occur at a frequency corresponding to the lower stability boundary of Fig. 2 while the unstable limit cycle will be at one of the higher frequencies of the upper stability boundary.

Finally, regions VI and VII represent totally unstable conditions for the case of rate-gyro threshold just as they do for a linear system since for these gains no realizable value of $N_r(\alpha_d)$ can lead to the stable region of Fig. 2. As shown in Fig. 5, the phase-plane sketch characterizing these regions consists of typical trajectories near an unstable singular point.

The preceding discussion has shown that the stability plot for the linear system (Fig. 2, $N_p = N_r = 1$) consists of only two separated regions which define the system as being either stable or unstable. The corresponding limit-cycle stability chart for the case of rate-gyro threshold, however, is divided into seven separated regions which represent five distinct types or degrees of stability.

Quantitative Limit-Cycle Data for Threshold

So far, only the nature of possible limit-cycle oscillations caused by threshold have been discussed. Now the amplitude and frequency of the oscillations will be quantitatively determined and compared with the results of analog computer tests.

The normalized amplitude α_d of the limit cycle can always be determined by computing the attenuation necessary for oscillations from Fig. 4 and then referring back to Fig. 3 to see what α_d value gives this attenuation. A more convenient and elucidating method, however, is to replot the portions of the stability chart and describing function of interest to a logarithmic scale; the describing-function graph can then be used as an overlay and the gain division indicated previously becomes simple graphical subtraction.

As an illustration, portions of Figs. 3 and 4 are replotted in terms of decibels (db) in Fig. 6. Here, the describing-function graph is drawn as $N_r(\alpha_d)$ versus N_p with α_d as a parameter; since the position gyro is presently assumed linear, N_p is identically unity (zero db) for all values of α_d. Two separate overlays of the describing function are shown on the stability chart of Fig. 6 as they would appear for the following gain conditions

Case A: $K_aK_r = 26$ db; $K_aK_p = 35.6$ db
Case B: $K_aK_r = 31$ db; $K_aK_p = 43$ db

For each of these system gains, the oscillation amplitude and frequency may be immediately determined by noting the values of α_d and ω corresponding to the intersections of the two curves. Note that for case B, both a stable and unstable solution exists for respectively the lower and upper intersection.

The accuracy with which the amplitude and frequency of limit-cycle oscillations can be predicted is illustrated in Fig. 7 which compares the results of analog computer studies and the proposed analytical technique. The analog setup employed a conventional biased-diode circuit for the threshold simulation; all harmonics from this circuit were allowed to circulate freely around the loop. The system rate gain was held at $K_aK_r = 20$ (26 db) and the position gain then varied to obtain the amplitude and frequency points and

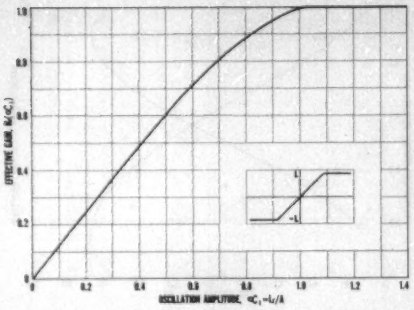


Fig. 8 (left). Describing function for limiting

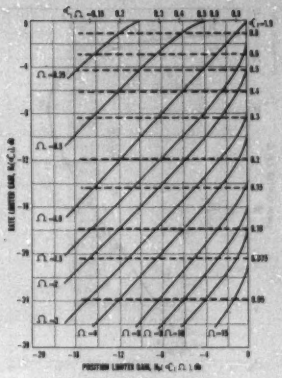


Fig. 9 (right). Describing - function overlay for simultaneous limiting in position and rate gyros. Each curve is associated with a single value of normalized frequency $\Omega = \gamma\omega$

curves shown in Fig. 7. The darkened triangles appearing on the analytically determined curves correspond to the similarly marked point in Fig. 6, case A.

It may be noted from Figs. 4 and 5 that the gain conditions $K_a K_r = 20$, $156 < K_a K_p < 167$ corresponds to region V wherein both stable and unstable oscillations are predicted. Both types were verified by the analog computer study and, as shown in Fig. 7, the proposed analytical method is capable of determining the amplitude and frequency of either type oscillation with quite suitable engineering accuracy. (It would appear, however, that the same reservations concerning higher harmonics applicable to conventional describing-function theory[3] are equally applicable to the new method.)

Rate-Gyro Limiting

The effects of limiting in the rate gyro can be determined by employing the same procedure just discussed for threshold after substituting the now-appropriate describing function for limiting shown in Fig. 8. Such an investigation reveals that the limit-cycle stability diagram remains divided into the same seven regions shown in Fig. 4. In all but region V, however, the stable and unstable oscillations shown in Fig. 5 become interchanged in going from threshold to limiting; in region V the frequencies of the two limit cycles become interchanged.

Analog-computer studies of the limit cycles resulting from rate-gyro limiting were also conducted. The analytical results were verified by this means to within essentially the same accuracy as that shown in Fig. 7 for threshold.

Investigation of the rate-gyro limiting case leads to a very interesting conclusion. The whole of region IV, Fig. 4, is plagued by a hard limit-cycle oscillation and yet this is the only region of practical operation from a linear standpoint. Since the threshold of a physical rate gyro usually increases in direct proportion to the maximum range, or limit value, it is usually not wise to specify a very large limiting value in an attempt to rationalize the limit cycle out of existence. What, then, can be done to circumvent this problem area? One answer is to deliberately introduce another nonlinearity into the system.

Rate- and Position-Gyro Limiting

It is obvious that position-gyro limiting would cause the describing-function loci of Fig. 6 to swing towards the left, for the value of N_p would then become less than unity. This suggests that operation in region IV, Fig. 4, might be possible without the potential hard limit cycle if position-gyro limiting were employed.

To investigate this possibility, assume that an oscillation in roll exists such that

$$\phi = A \sin \omega t \tag{8}$$

$$\phi = \int \phi dt = -\frac{A}{\omega} \cos \omega t \tag{9}$$

If now the rate signal is limited to the value $\pm L_r$, and the position signal is limited to $\pm L_p$, the describing function $N(\alpha)$ for each limiter will be as shown in Fig. 8; here, α is defined for the rate limiter as

$$\alpha = \frac{L_r}{A} = \alpha_1 \tag{10}$$

and for the position limiter,

$$\alpha = \frac{L_p}{A/\omega} = \frac{L_r}{A}\frac{L_p}{L_r}\omega = \alpha_1\gamma\omega = \alpha_1\Omega \tag{11}$$

From equations 10 and 11, a new overlay chart of $N_r(\alpha_1)$ versus $N_p(\alpha_1\Omega)$ may be constructed with loci for several values of normalized frequency, $\Omega = \gamma\omega$, as shown in Fig. 9. This overlay can be used in conjunction with a linearized-gain stability plot just as was done in Fig. 6 except now only those intersections occurring at like frequencies are valid. If a region can be found in which no valid intersections occur, then no limit cycles due to gyro limiting are possible in that region.

The determination of such a region is demonstrated in Fig. 10 which shows a few values of the $N_r(\alpha_1)$ versus $N_p(\alpha_1\Omega)$ overlay drawn on the stability chart for the special case of $\gamma = L_p/L_r = 1/5$ second.

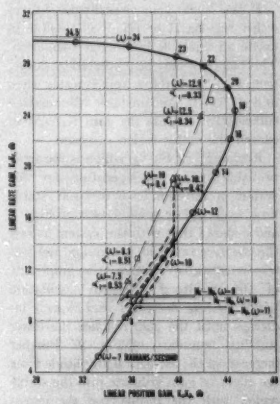


Fig. 10. Logarithmic plot of stability diagram and describing-function overlay for simultaneous gyro limiting in the ratio of $\gamma = 1/5$

- – – – Separates regions of stable and hard limit-cycle operation
- △ Boundary solutions obtained by describing function
- □ Solutions obtained from analog computer

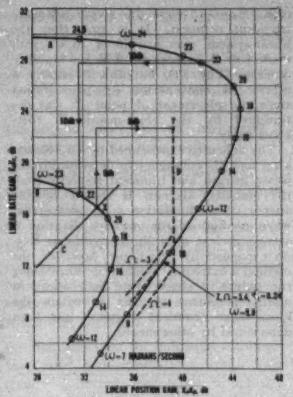


Fig. 11. Logarithmic plot of stability diagram showing synthesis procedure

Note that if the describing-function overlay were moved up from the gain value shown, no more valid (like-frequency) intersections would occur. But if the gain were lowered, two proper intersections would result: one corresponding to a hard, unstable limit cycle of frequency $\omega = 10.6$ radians per second, and the other a stable oscillation of somewhat greater amplitude and lower frequency. It is now apparent that as the system gain approaches a value for which valid intersections are only marginally possible, the two limit cycles coalesce to destroy one another.[4] Such a gain, therefore, represents one point on a boundary line which separates the stability chart into regions of stable and potentially unstable (hard limit cycle) operation.

The heavy dashed line of Fig. 10 shows the analytically determined locus of system gains resulting in neutral limit-cycle stability for $\gamma = 1/5$ second. Also shown for comparison are the amplitude and frequency of several of these marginal oscillations as determined both analytically and by the analog computer. Note the excellent agreement in all respects.

A Synthesis Procedure

Having developed the tools necessary to handle the nonlinearities in the system, suppose now that it is desired to determine the best gains and gyro limit values to satisfy the following design requirements:

1. The feedback gain ratio shall be adjusted such that $K_r/K_p = 0.15$ second.
2. The linear-system gain margin with respect to K_a shall be 10 db.
3. The hard limit-cycle gain margin with respect to K_a shall be 6 db.
4. The system shall be capable of measuring roll angles up to ± 25 degrees.
5. The amplitude of soft, limit-cycle oscillations due to rate-gyro threshold shall be held to a minimum.

The first requirement is based on a linear root locus study in which the amount of rate-to-position feedback required to provide an adequate transient response in the linear mode of operation was determined.

In the second and third requirements, the gain margin is expressed in terms of K_a since, as mentioned previously, this gain is subject to rather wide and unpredictable variations. The limit-cycle gain margin is specified somewhat lower than that for the linear case since not only the system gain but also the level of excitation existing simultaneously must be favorable before oscillations can develop. The relative weighting of the two gain margins is largely a matter of judgement.

The fourth and fifth requirements state, in essence, that the position-gyro limit be set at 25 degrees. For if this limit were higher than necessary, the rate-gyro limiter would also have to be higher to provide a given limit-ratio value γ, and this in turn would increase the threshold level of the rate gyro because of physical construction considerations.

The synthesis procedure is shown in Fig. 11. First, the linear frequency locus is plotted as curve A. Next, curve B is established by projecting each point of curve A back and down 10 db; obviously curve B represents a boundary outside of which the linearized gain margin with respect to K_a is less than 10 db. The condition for $K_r/K_p = 0.15$ is depicted as the straight line C. Thus the point X satisfies the first two requirements previously listed and represents the desired gain condition.

By virtue of requirement 3, the point Y may be determined as falling on the limit-cycle stability boundary; therefore the describing-function overlay, curve D, must furnish a marginally stable solution when its zero-zero db point is aligned with Y. This marginally stable solution is shown as point Z in Fig. 11 and occurs for $\omega = 9.8$ radians per second and $\Omega = \gamma\omega = 3.4$. From these data, the limit ratio γ may be solved for as

$$\gamma = \frac{L_p}{L_r} = \frac{\Omega}{\omega} = \frac{3.4}{9.8} \text{ second} \qquad (12)$$

Now from requirement 4 and equation 12, the desired value of rate limit, L_r, becomes

$$L_r = \frac{L_p}{\gamma} = \frac{(25)(9.8)}{3.4} = 72 \text{ degrees per second} \qquad (13)$$

In summary, the system which meets the specified requirements has the following parameters:

$K_a K_r = 16.6$ db $= 6.8$ per second
$K_a K_p = 33.2$ db $= 46$ per second squared
$L_p = 25$ degrees
$L_r = 72$ degrees per second

Conclusions

The new describing-function technique presented herein appears to offer quite substantial advantages in interpretation and flexibility over the more conventional approach to limit-cycle analysis when the system nonlinearities are of the gain-controlling type. A slight modification[3] would allow phase-shifting nonlinearities to be handled also but with dubious benefits.

The additional utility of the new method stems from having two generalized system parameters with which to work rather than the single gain associated with open-loop diagrams.[2] The price paid for this convenience is the somewhat laborious derivation and computation of the characteristic equation of the closed-loop system, particularly if the order of the equation is high. To obtain the same amount of information and insight by the conventional method would, however, lead to vastly more labor since families of open-loop frequency loci would be necessary to account for the second parameter.

For the system analyzed in this paper, it was convenient to choose two gains for the co-ordinates of the stability diagram. In other problems it might be more appropriate to choose one gain and some other system parameter,[1] a time constant for example. In other words, the method is considerably more flexible than is demonstrated by the particular analysis carried out herein.

References

1. A New Stability Criterion of Linear Servomechanisms by a Graphical Method, **Toshio Numakura, Takeo Miura.** *AIEE Transactions*, vol. 76, pt. II (*Applications and Industry*), Mar. 1957, pp. 40–48.

2. Analyses of Servomechanisms by Describing Function Method on a Plane with Parameters as Coordinates, **Toshio Numakura, Takeo Miura, Akira Kamei.** *AIEE CP57-1033* (available on request).

3. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, **E. Calvin Johnson.** *AIEE Transactions*, vol. 71, pt. II (*Applications and Industry*), July 1952, pp. 169–81.

4. Non-Linear Mechanics (book), **N. Minorsky.** J. W. Edwards, Inc., Ann Arbor, Mich., 1947, pp. 68–73.

Discussion

Toshio Numakura (Hitachi Central Research Laboratory, Kokubunji, Tokyo, Japan): The author has solved a complicated system which contains several nonlinearities by using the stability diagrams introduced in references 1 and 2 of the paper.

The new ideas added to the describing-function method and the works cited are very interesting.

Some problems may remain unsolved in the last section of the paper, "A Synthesis Procedure," in which the author intends to predict the transient response of the nonlinear system by use of the stability diagram. Fig. 11 does not provide any indication of the phase margin or slope of the gain characteristic as does Bode's diagram, so it is not possible to predict the transient response exactly. Though the discusser and his colleagues intended to find a suitable method by which the transient response could be predicted, they failed to find any not only in the case of the nonlinear system but for the linear system as well. Therefore the optimum parameter conditions should be found by using the stability diagram in conjunction with an analog computer.

H. J. Harrington: The author appreciates the comments of Dr. Numakura and the resulting opportunity to elaborate on the paper.

Actually, there is no intent in the paper to predict the transient response by means of the stability diagram. Thus, in the section referred to by the discusser, it is pointed out that a complementary root-locus analysis furnished the gain ratio value of $K_r/K_p = 0.15$ required to give an adequate transient response. It is true that the linear-system gain margin of 10 db with respect to K_a is marked off on the stability diagram (Fig. 11), but this is mostly a matter of convenience for presentation since the placement of point X was determined by the afore-mentioned root-locus study.

The difficulty experienced by Dr. Numakura and his colleagues in determining the phase margin and transient response from the stability diagram is shared by the author. This limitation is so severe that when used alone the stability diagram is almost useless as a linear synthesis tool. As expanded by this paper, however, the method is extremely valuable in establishing the nature of steady-state limit cycles in mildly nonlinear systems in which the linear-range transient response has been determined by other means.

New Methods of Simplifying Boolean Functions

R. L. HOWARD
ASSOCIATE MEMBER AIEE

ECONOMY IS ONE of the most important considerations in the design of any system. Maximum savings can be achieved in the design of switching circuits by performing the desired logic in the simplest minimum form applicable to the equipment utilized. Boolean algebra is often used to express the functions of switching circuits and these Boolean expressions generally appear in the final form of a sum of products or a product of sums. The form that presents the most economical application is determined by the equipment. It is therefore necessary to be able to transform and simplify Boolean functions so that they will appear in the predetermined form that will yield the maximum savings.

Methods for performing the simplification and transformation of Boolean functions have been presented in previous publications by Karnaugh,[1] Quine,[2] and McCluskey.[3] Each of these processes becomes extremely complicated for complex functions. Furthermore, some duplications are encountered which repeat information during the process that initially had been stated in the original function.

The following discussion describes a new procedure of simplifying and transforming any Boolean function without some of the restrictions encountered in previously presented procedures. The original Boolean function may consist of any number of variables and may be expressed in any form (sum of product, product of sums or a combination of both). Without changing the form of the original function the same basic rules apply when either simplification and transformation is made; only the recording of the original function is changed. This provides a systematic process for the reduction of static or conventional control systems and also an easily programmed process for digital computer application.

Discussion

For clarity the explanation of the method includes a sample Boolean function. Wherever the sample is concerned in the explanation that portion will be labelled "sample." Also included wherever possible, are examples which are independent of the sample function. Each of these will be labelled "Example."

The sample Boolean function is

$$X = A \cdot B[C \cdot D \cdot E + A \cdot \bar{C}\,\bar{E}(C+D)] = AB[CDE + A\bar{C}\bar{E}(C+D)]$$

where

$+$ = OR function
$\cdot$ = AND function

The "variables" of a function are the number of different letters concerned. The "literals" of a function are the individual entries in the function. Thus, the sample function contains five variables and ten literals. Fig. 1 shows logic diagrams of the sample function.

The Basic Procedure of the Method

Preparation of the Function for Transformation and Simplification

1. Determine the number of variables contained in the original expression.

Sample: There are five variables contained in the expression (A, B, C, D and E).

2. Assign a digit to each variable.

Digits
Sample: $E\ D\ C\ B\ A$

3. The digital representation of each product is now written as follows:

If the literal indicates the presence of the variable, a "1" will be placed in the corresponding digit for that term.

If the literal indicates the absence (negation) of the variable, a "0" will be placed in the corresponding digit for that term.

If the literal does not appear in the term a dash (–) will be placed in the corresponding digit for that term.

Sample: $AB[CDE + A\bar{C}\bar{E}(C+D)]$

$E\ D\ C\ B\ A$ $\quad$ $E\ D\ C\ B\ A$
$-\ -\ -\ 1\ 1$ $\quad$ $[1\ 1\ 1\ -\ -+$

$E\ D\ C\ B\ A$ $\quad$ $E\ D\ C\ B\ A$
$0\ -\ 0\ -\ 1$ $\quad$ $(-\ -\ 1\ -\ -+$

$E\ D\ C\ B\ A$
$-\ 1\ -\ -\ -)]$

4. For the conversion of the original

Paper 59-91, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted November 3, 1958; made available for printing December 16, 1958.

R. L. HOWARD is with Westinghouse Electric Corporation, Pittsburgh, Pa.

expression into the sum of products (which will be referred to as the AND form): (*a*). Wherever an AND function is indicated in the original expression the two terms it concerns will be combined. (*b*). Wherever an OR function is indicated in the original expression the two terms it concerns will not be combined.

5. For the conversion of the original expression into the product of sums (which will be referred to as the OR form): (*a*). Wherever an OR function is indicated in the original expression the two terms it concerns will be combined. (*b*). Wherever an AND function is indicated in the original expression the two terms it concerns will not be combined.

6. Symbols used for indicated combinations and noncombinations are a long dash (—) between two terms to indicate that they are to be combined; and a long marked dash (-+-) used between two terms to indicate that they are not to be combined. These symbols will vary in length to correspond to the original expression.

7. The original expression is recorded (rewritten). The digit representation of each term is written, one term beneath the other. The proper indications for combination or noncombination are included in the following vertical representation of the expression. Note that the digital representation of the expression remains the same for both transformations. The only difference is that the combination and noncombination indications are reversed.

Sample:

```
E D C B A    E D C B A
- - - 1 1  [1 1 1 - -+
    E D C B A    E D C B A
    0 - 0 - 1  (- - 1 - -+
                E D C B A
                - 1 - - -)]
```

To AND Form

```
E D C B A
- - - 1 -
---------
- - - - 1
---------
1 - - - -
---------
- 1 - - -
---------
- - 1 - -
---------
0 - - - -
---------
- - 0 - -
---------
- - - - 1
---------
- - 1 - -
---------
- 1 - - -
```

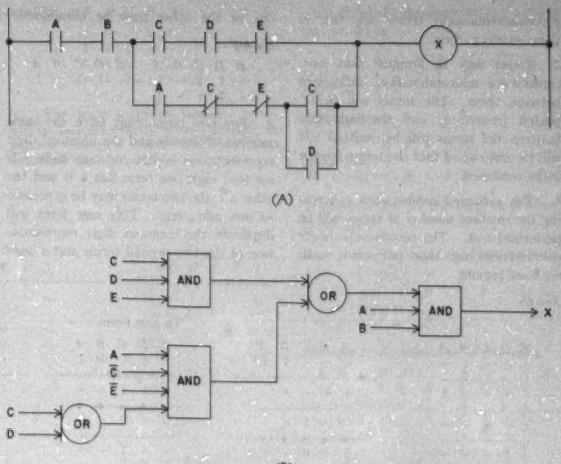


Fig. 1. Sample expressions

A—Relay diagram
B—Static diagram

To OR Form

```
E D C B A
- - - 1 -
---------
- - - - 1
---------
1 - - - -
---------
- 1 - - -
---------
- - 1 - -
---------
0 - - - -
---------
- - 0 - -
---------
- - - - 1
---------
- - 1 - -
---------
- 1 - - -
```

Rules for Transforming the Function

1. Rules for any digit of two terms that are to be combined are:

(*a*). A 0 in the digit of one term and a 1 in the corresponding digit of the other term. The two terms cannot be combined and there will be no resulting term.

Example: Digit $\frac{1}{0}$ = no resulting term

(*b*). A 0 in the digit of one term and a 0 in the corresponding digit of the other term will yield a 0 in the same digit of the resulting term.

Example: Digit $\frac{0}{0}=0$

(*c*). A 0 in the digit of one term and a dash (–) in the corresponding digit of the other term will yield a 0 in the same digit of the resulting term.

Example: Digit $\frac{0}{-}=0$

(*d*). A 1 in a digit of one term and a 1 in the corresponding digit of the other term will yield a 1 in the same digit of the resulting term.

Example: Digit $\frac{1}{1}=1$

(*e*). A 1 in a digit of one term and a dash (–) in the corresponding digit of the other term will yield a 1 in the same digit of the resulting term.

Example: Digit $\frac{1}{-}=1$

(*f*). A dash (–) in a digit of one term and a dash (–) in the other term will yield a dash (–) in the same digit of the resulting term.

Example: Digit $\frac{-}{-}=-$

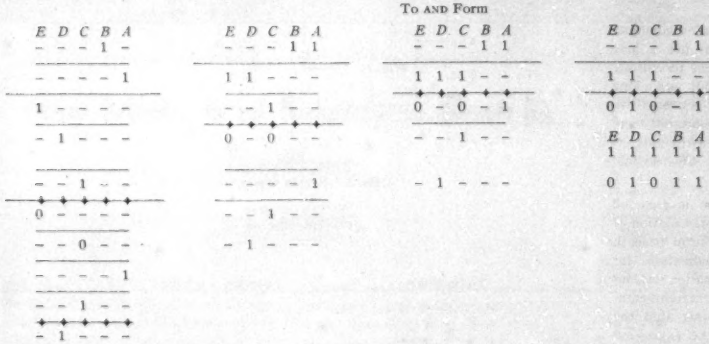
2. Noncombinable terms are left in their existing form.

3. Terms may be grouped that have consecutive noncombination indications between them. The terms will be repeated (rewritten) and the indication between the terms will be omitted. It will be understood that the terms are not to be combined.

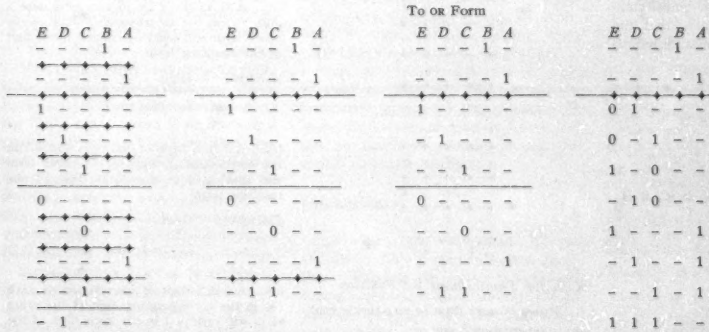
4. The indicated combination concerning the smallest number of terms will be performed first. The successively larger combinations are then performed until no more remain.

Simple:

To AND Form



To OR Form



Rules for Simplifying the Functions

A literal will be considered to be contained in a digit if that digit has a 0 or 1 in it. The literal digit representation of a term is the 0 and/or 1 configuration contained in its digits. If a dash (–) is present in the digit the literal is not contained in that digit.

1. If two identical terms are repeated, one or the other may be eliminated.

Example:

E	D	C	B	A		E	D	C	B	A
–	1	1	–	1	=	–	1	1	–	1
–	1	1	–	1						

2. For two terms that have the same number of literals and the identical digit representation except for one digit. If for that digit one term has a 0 and the other a 1, the two terms may be expressed as one new term. This new term will duplicate the common digit representation of the two orginal terms and a dash (–) will be placed in the digit that varied from 0 to 1 (or 1 to 0).

Example:

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	=	–	1	–	–	1
–	1	–	0	1						

3. For two terms, one of which contains a lesser number of literals than the other and also has its literal digit representation duplicated in the other term. The term with the greater number of literals is redundant and may be eliminated.

Example:

E	D	C	B	A		E	D	C	B	A
–	1	0	1	–	=	–	1	0	1	–
–	1	0	1	1						

4. For two terms, one of which contains a lesser number of literals than the other and also has its literal digit representation duplicated except for one digit in the other term, if for that one digit one term has a 0 and the other a 1, the term containing the least number of literals is retained. The other term becomes a new term that duplicates the common digit representation of the two original terms and a dash (–) is placed in

the digit that varied from a 0 to 1 (or 1 to 0).

Example:

E	D	C	B	A		E	D	C	B	A
–	1	1	–	1	=	–	1	1	–	1
–	1	0	1	1		–	1	–	1	1

The sample for the AND form used in item *4* under "Rules for Transforming the Function," indicates that the remaining two terms cannot be simplified and therefore constitute the final expression.

Sample:

E	D	C	B	A	
–	–	–	1	1	*P*
1	1	1	–	–	*P*
0	–	0	–	1	*P*
–	1	1	–	–	*S*

(*c*). The logic functions between terms will be handled exactly as was outlined in items *4* through *7* under "The Basic Procedure of the Method."

To illustrate the combination consider two products

	E	D	C	B	A	
$D{\cdot}B{\cdot}A$ =	–	1	–	1	1	*P*
$\bar{E}{\cdot}D{\cdot}B$ =	0	1	–	1	–	*P*

and two sums

	E	D	C	B	A	
$(\bar{E}+B+\bar{A})$ =	0	–	–	1	0	*S*
$(\bar{D}+C+B)$ =	–	0	1	1	–	*S*

Transformation to the sum of products (AND) form:

Example 1. Combination of two products

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	*P*	0	1	–	1	1
0	1	–	1	–	*P*					

Example 2. Combination of two sums

E	D	C	B	A		E	D	C	B	A
0	–	–	1	0	*S*	0	0	–	–	–
–	0	1	1	–	*S*		0	–	1	–
						0	–	–	1	–
							0	–	1	–
						–	–	1	1	–
						–	–	–	1	–
						–	0	–	–	0
						–	–	1	–	0
						–	–	–	1	0

Example 3. Combination of a product AND sum

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	*P*	0	1	–	1	1
0	–	–	1	0	*S*	–	1	–	1	1

Transformation to the product of sum (OR) form:

Example 4. Combination of two products

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	*P*	0	1	–	–	–
0	1	–	1	–	*P*	–	1	–	–	–
						0	–	–	1	–
						–	1	–	1	–
						–	–	–	1	–
						0	–	–	–	1
						–	1	–	–	1
						–	–	–	1	1

Example 5. Combination of two sums

E	D	C	B	A		E	D	C	B	A
0	–	–	1	0	*S*	0	0	1	1	0
–	0	1	1	–	*S*					

Example 6. Combination of a product OR a sum OR a product

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	*P*	0	1	1	0	–
0	–	1	0	–	*S*	0	–	1	0	1
0	1	–	1	–	*P*	0	1	–	1	–

E	D	C	B	A
0	1	1	0	–
0	–	1	0	1
0	1	1	0	–
0	1	1	0	1

To OR Form

E	D	C	B	A		E	D	C	B	A		E	D	C	B	A
(2) –	–	–	1	–	(2) –	–	–	1	–		(2) –	–	–	1	–
(3) –	–	–	–	1	(4) 0	1	–	–	–		(4) 0	1	–	–	–
(4) 0	1	–	–	–	(5) 0	–	1	–	–		(5) 0	–	1	–	–
(5) 0	–	1	–	–	(6) 1	–	0	–	–		(6) 1	–	0	–	–
(6) 1	–	0	–	–	(7) –	1	0	–	–		(7) –	1	0	–	–
(7) –	1	0	–	–	(3)(8)(9)(10)	–	–	–	1		(3)(8)(9)(10)	–	–	–	1
(8) 1	–	–	–	1	(11) 1	1	1	–	–		(11)(12)(13) –	1	1	–	–
(9) –	1	–	–	1	(12)(13) –	1	1	–	–						
(10) –	–	1	–	1											
(11) 1	1	1	–	–											
(12) –	1	1	–	–											
(13) –	1	1	–	–											

E	D	C	B	A		E	D	C	B	A
(2) –	–	–	1	–	(2) –	–	–	1	–	
(4) 0	1	–	–	–	(5) 0	–	1	–	–	
(5) 0	–	1	–	–	(6) 1	–	0	–	–	
(6) 1	–	0	–	–	(3)(8)(9)(10)	–	–	–	1	
(3)(8)(9)(10)	–	–	–	1	(4)(7)(11) –	1	–	–	–	
(7)(11)(12)(13) –	1	–	–	–	(12)(13)					

Terms 3, 8, 9 and 10 are simplified by rule *3*.
Terms 12 and 13 are simplified by rule *1*.
Terms 11 and 12, and 13 are simplified by rule *3*.
Terms 7 and 11, 12, and 13 are simplified by rule *2*.
Terms 4 and 7, 11, 12, and 13 are simplified by rule *3*.

Reductions in the Basic Procedures of the Method

It is possible to shorten the basic procedure which has been outlined in detail. This reduction is accomplished by retaining the basic rules of combination and simplification and by adding the following modifications.

1. In order to record the original function for simplification and transformation:

(*a*). The same procedure is followed as outlined in items *1*, *2*, and *3* under "The Basic Procedure of the Method" except that the digital representation of each term instead of each product will be written.

Sample: $AB[CDE+A\bar{C}\bar{E}(C+D)]$

E	D	C	B	A		E	D	C	B	A
(–	–	–	1	1)	[1	1	1	–	–	+]
0	–	0	–	1	(–	1	1	–	–)]	

(*b*). To record the function the digital representation of each term is written, one term beneath the other. A *P* will be placed next to all terms that are products. An *S* will be placed next to all terms that are sums.

Sample: To AND Form

E	D	C	B	A	
–	–	–	1	1	*P*
1	1	1	–	–	*P*
0	–	0	–	1	*P*
–	1	1	–	–	*S*

To OR Form

E	D	C	B	A	
–	–	–	1	1	*P*
1	1	1	–	–	*P*
0	–	0	–	1	*P*
–	1	1	–	–	*S*

2. Combination. The basic rules are applied for indicated combinations with these additions for transformation to the AND form: (*a*). A term that is a product will be combined as a whole term; (*b*). A term that is a sum (*S*) will be combined digit by digit.

For transformation to the OR form: (*a*). A term that is a sum will be combined as a whole term; (*b*). A term that is a product (*P*) will be combined digit by digit.

Upon close inspection of these examples it appears that some of the resulting terms of an indicated combination do not appear. This is due to part (*a*), item *1* under "Rules for Transforming the Function."

3. Simplification. The examples in the preceding section were not simplified. Further inspection of example *2* yields that one term contains only the *B* signal. By rule *3*, all other terms that contain *B* (a 1 in the *B* digit) are redundant and may be eliminated. This simplification may be accomplished at the same time the combination is made, thereby directly reducing the number of resulting terms. To accomplish this, the following rule may be used:

If in performing the indicated combination (for either transformation) of literals or whole terms, one of the items has its digital representation duplicated in the resulting term, all other combinations of this item will prove redundant and therefore need not be considered.

Example (from Exaınple *2*)

	E	D	C	B	A	
	0	–	–	1	0	*S*
	–	0	1	1	–	*S*

	E	D	C	B	A
	–	–	–	1	–
	0	0	–	–	–
	0	–	1	–	–
	–	0	–	–	0
	–	–	1	–	0

Example (from Example *6*)

	E	D	C	B	A	
(2)	–	1	–	1	1	*P*
(3)	0	–	1	0	–	*S*
(4)	0	1	–	1	–	*P*

	E	D	C	B	A
(2)(3)	0	1	1	0	–
	0	–	1	0	1
(4)	0	1	–	1	– *P*

	E	D	C	B	A
(2)(3)(4)	0	1	1	0	–
	0	–	1	0	1

Sample:

To AND Form

(1)

	E	D	C	B	A	
(2)	–	–	–	1	1	*P*
(3)	1	1	1	–	–	*P*
(4)	0	–	0	–	1	*P*
(5)	–	1	1	–	–	*S*

(2)

	E	D	C	B	A	
	–	–	–	1	1	*P*
(3)	1	1	1	–	–	*P*
(4)(5)	0	1	0	–	1	

(3)

	E	D	C	B	A
	1	1	1	1	1
	0	1	0	1	1

To OR Form

(1)

	E	D	C	B	A	
(2)	–	–	–	1	1	*P*
(3)	1	1	1	–	–	*P*
(4)	0	–	0	–	1	*P*
(5)	–	1	1	–	–	*S*

(2)

	E	D	C	B	A	
(2)	–	–	–	1	1	*P*
	1	–	0	–	–	
	1	–	–	–	1	
(3)(4)(5)	0	1	–	–	–	
	–	1	0	–	–	
	–	1	–	–	1	
	0	–	1	–	–	
	–	–	1	–	1	
	–	1	1	–	–	

(3)

	E	D	C	B	A
a	–	–	–	–	1
b	–	–	–	1	–
c	1	–	0	–	–
d	1	–	–	–	1
(2)(3)(4)(5) *e*	0	1	–	–	–
f	–	1	0	–	–
g	–	1	–	–	1
h	0	–	1	–	–
i	–	–	1	–	1
j	–	1	1	–	–

(4)

	E	D	C	B	A
a	–	–	–	–	1
b	–	–	–	1	–
c	1	–	0	–	–
(2)(3)(4)(5) *e*	0	1	–	–	–
f	–	1	0	–	–
h	0	–	1	–	–
j	–	1	1	–	–

(5)

	E	D	C	B	A
a	–	–	–	–	1
b	–	–	–	1	–
c	1	–	0	–	–
e	0	1	–	–	–
h	0	–	1	–	–
fj	–	1	–	–	–

(6)

	E	D	C	B	A
a	–	–	–	–	1
b	–	–	–	1	–
c	1	–	0	–	–
h	0	–	1	–	–
fj	–	1	–	–	–

Obtaining the Minimum Form of a Solution

After all indicated combinations have been performed the resulting terms constitute a solution. This solution is not necessarily in the minimum form and the following process may be used to obtain a minimum form of a solution. The process will indicate that a term is redundant if its function is provided for in two other terms. In order for this to be valid the two terms must contain fewer or an equal amount of literals than the third term. The process is used as follows:

1. Consider the terms containing the greatest number of literals first.

(*a*). Substitute a 0 and a 1 in a digit containing a dash (–) in the considered term. This will yield two new forms of the considered term.

(*b*). Compare one of these new forms to the remaining terms. If it duplicates the literal digit representation of another term the other new form of the considered term will be compared. If this form duplicates the literal digit representation of another term the considered term is redundant and will be eliminated.

(*c*). Should the first 0,1 substitution not prove the term redundant the procedure will be repeated for another digit containing a dash in the considered term. This will continue until the term is determined redundant or that it is to be maintained.

2. The next term containing the greatest number of literals is then considered and checked for redundancy in the same manner. This is done until there are no more terms to give the necessary comparison, i.e., that there are not three terms that have at least an equal amount of literals.

The sample function for both transformations is already in the minimum form.

The following examples will illustrate the procedure.

Example 7:

	C	B	A
(2)	–	1	1
(3)	1	–	0
(4)	1	1	–

	C	B	A
(2)	–	1	1
(3)	1	–	0

All terms have the same number of literals so that each will be considered.

Term (*2*)

Making the 0,1 substitution in the *C* digit yields the new forms of 0 1 1 and 1 1 1.

Comparing 0 1 1 to terms (3) and (4) shows that neither term has its literal digit representation duplicated. Therefore term (2) is not redundant and must be retained.

Term (3)

Making the 0,1 substitution in the B digit yields the new forms 1 0 0 and 1 1 0.

Comparing the 1 1 0 form with terms (2) and (4) it duplicates the literal digit representation of term (4). The 1 1 0 form is then compared to term (2) and it does not duplicate this term's literal digit representation. Therefore term (3) is not redundant and must be retained.

Term (4)

Making the 0,1 substitution in the A digit yields the new forms 1 1 0 and 1 1 1.

Comparing the 1 1 0 form to terms (2) and (3) it duplicates the literal digit representation of term (3). The 1 1 1 form is then compared to term (2) and it duplicates the literal digit representation of term (2). Therefore term (4) is redundant and is eliminated.

Example 8:

	E	D	C	B	A
(2)	–	0	–	0	1
(3)	–	–	1	1	1
(4)	–	1	–	1	–
(5)	–	0	1	–	–
(6)	0	–	0	0	1
(7)	0	1	0	–	–

	E	D	C	B	A
(2)	–	0	–	0	1
(4)	–	1	–	1	–
(5)	–	0	1	–	–
(7)	0	1	0	–	–

Substitution of a 0, 1 in the D digit of term (6) yields the new forms which duplicate the literal digit representation of term (2) and (7). Therefore term (6) is redundant and is eliminated.

Substitution of 0,1 in the D digit of term (3) yields new forms which duplicate the literal digit representation of terms (4) and (5). Therefore term (3) is redundant and is eliminated.

The other terms are checked for redundancies and none are found. Therefore all must be retained.

After becoming familiar with this process the substitution and comparisons will be made mentally and the redundant terms eliminated by inspection.

This process allows a term that is provided in two other terms to be eliminated. Elaboration of the process so that a function that is provided in three or more terms can be eliminated. Generally the elimination of these redundant terms will yield the minimum form of the solution. Many other methods (including those outlined in the Karnaugh, Quine, and McCluskey publications) may be used to obtain the minimum form of the solution.

Writing the Final Expression

After all indicated combinations have been performed and all simplifications have been made the remaining terms are in a group. For the conversion to the AND form each remaining term will be a product and the terms will be separated by OR functions. For the conversion to the OR form each remaining term will be a sum and the terms will be separated by AND functions.

For either conversion each term will be written as follows:

1. The variable will be present if a 1 appears in its digit of the particular term.
2. The negation of the variable will be written if a 0 appears in its digit of the particular term.
3. The variable will not be written for a particular term if a dash (–) appears in its digit.

Sample:

To AND Form

E	D	C	B	A	
1	1	1	1	1	$A\ B\ C\ D\ E$
0	1	0	1	1	$A\ B\ \bar{C}\ D\ E$

Final expression: $X = ABCDE + AB\bar{C}DE$

To OR Form

E	D	C	B	A	
–	–	–	1	–	B
0	–	1	–	–	$(\bar{C}+E)$
1	–	0	–	–	$(C+E)$
–	–	–	–	1	A
–	1	–	–	–	D

Final expression: $X = B(C+E)(\bar{C}+E)AD$
$= ABD(C+E)(\bar{C}+E)$

Figs. 2(A) and (B) show relay and static logic diagrams of the sample expression for both transformations.

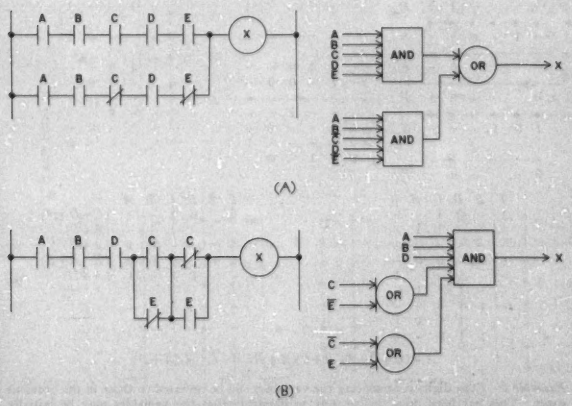


Fig. 2. Relay and static diagrams of transformation of sample expression

A—AND form
B—OR form

Summary and Conclusions

In summation, two new methods of simplification and transformation of Boolean functions have been presented in this paper. These methods are utilized to obtain a minimum solution for a desired form and may be referred to as conversions of the original function. One method is used to convert the original function to a minimum sum of products and the other method is used to convert the original function to a minimum product of sums. The two methods parallel each other in that they utilize the same basic rules, the major difference being the recording of the original expression. Because of this similarity the two methods have been explained as one basic procedure. The fundamental principles of this basic procedure have been outlined in detail in the discussion. A reduction has also been explained that allows the basic procedure to be considerably shortened. A "sample" Boolean function and independent "examples" have been included to better explain the procedure. Examples of conversion to both forms of two Boolean functions have also been included in the discussion.

The following advantages occur through the use of the procedure presented in this paper:

1. The method for either conversion is based upon the postulates and theorems of Boolean algebra. Therefore, (a) any extension of the procedure or any portion of a

similar process that is also based on Boolean algebra may be incorporated to expedite the procedure; and (*b*) the manipulation of the original expression in lettered variable form by standard Boolean algebra methods is eliminated.

2. The original Boolean function appearing in either pure form (sum of products or product of sums) or a combination of both forms may be converted directly. Therefore, (*a*) it is not necessary that the original function be initially presented in either pure form to accomplish minimization or conversion, as it is in the Karnaugh map method; and (*b*) it is not necessary to analyze the original expression for the basic transmission functions as it is in McCluskey's method. This eliminates duplication of information stated initially in the original function. This duplication is encountered in the combining of these basic transmission functions to achieve the minimum form of a solution. For either conversion by the new methods the original function is the true base.

3. No restriction is placed on the number of variables or the complexity of the Boolean function for either conversion.

4. The basic procedure is the same for either conversion. Therefore by becoming acquainted with one of the new methods of conversion it can be applied easily to accomplish the other, eliminating the necessity of learning two distinctly different methods.

5. Although not a new concept the 1,0 and dash (–) variable representation utilized presents first, a clear, concise and easily handled representation of the original function for either conversion; second, a step-by-step tabular form for either conversion that will reduce the possibility of omitting literals of terms or whole terms; and third, a record that can be checked conveniently for errors.

6. The procedure is such that for either conversion most of the redundant terms are disregarded or eliminated as the conversion progresses. This means a minimum number of redundant terms are carried through to the last step.

In conclusion, from the fundamental precepts of the procedure presented:

1. New parallel methods of simplifying (obtaining a minimum form) Boolean functions have been provided for static and conventional relay-circuitry reduction. The equipment itself determines the conversion necessary to provide for maximum savings.

2. The use of standard Boolean algebra operation on the function has been eliminated in these new methods. Some of the restrictions encountered in using other methods (previously mentioned) have also been eliminated.

3. The rigidly systematic parallel methods and tabular 1,0 and dash (–) variable representation present an easily programmed process for digital computer application.

Extension of these basic ideas will surely provide even further reduction of the procedure.

Appendix

The preceding sample and examples considered no more than five variables. The following examples illustrate how both transformations may be accomplished for expressions containing more than five variables. The method utilized in these examples is the reduction of the basic procedure.

Example 1: $(C+\bar{D}+BF)[AB(\bar{C}+E)+\bar{B}D]+A\bar{B}C[(D+E+F)+\bar{B}\bar{F}]$

To AND Form

F	E	D	C	B	A	
–	–	–	1	–	–	P
–	–	0	–	–	–	P
1	–	–	–	1	–	P
–	–	–	–	1	1	P
–	1	–	0	–	–	S
–	–	1	–	0	–	P
–	–	–	1	0	1	P
1	0	1	–	–	–	S
0	–	–	–	0	–	P

F	E	D	C	B	A
–	–	–	1	–	–
–	–	0	–	–	–
1	–	–	–	1	–
–	1	–	–	1	1
–	–	–	0	1	1
–	–	1	–	0	–
1	–	–	1	0	1
–	0	–	1	0	1
0	–	–	1	0	1
–	–	1	1	0	1

F	E	D	C	B	A	
–	–	–	1	–	–	
–	–	0	–	–	–	
1	–	–	–	1	–	
–	1	–	–	1	1	
–	–	–	0	1	1	
–	–	1	–	0	–	
–	–	–	1	0	1	
–	0	–	1	0	1	
–	–	1	1	0	1	

F	E	D	C	B	A	
–	1	–	1	1	1	
–	–	1	1	0	–	
–	–	0	0	1	1	
1	–	–	0	1	1	
–	–	–	1	0	1	

F	E	D	C	B	A	
–	1	–	1	–	1	ACE
–	–	1	1	0	–	$\bar{B}CD$
–	–	0	0	1	1	$AB\bar{C}\bar{D}$
1	–	–	0	1	1	$AB\bar{C}F$
–	–	–	1	0	1	$A\bar{B}C$

$$X=ACE+\bar{B}CD+AB\bar{C}\bar{D}+AB\bar{C}F+A\bar{B}C$$

To OR Form

F	E	D	C	B	A	
–	–	–	1	–	–	S
–	–	0	–	–	–	S
1	–	–	–	1	–	P
–	–	–	–	1	1	P
–	1	–	0	–	–	S
–	–	1	–	0	–	P
–	–	–	1	0	1	P
1	0	1	–	–	–	S
0	–	–	–	0	–	P

F	E	D	C	B	A
1	–	0	1	–	–
–	–	0	1	1	–
–	–	1	–	1	–
–	–	1	–	–	1
–	–	–	–	0	1
–	1	–	0	0	–
–	–	1	–	–	1
–	–	1	1	1	–
–	1	1	0	0	–
–	–	1	–	1	1
–	1	1	0	–	1

F	E	D	C	B	A
1	–	0	1	–	–
–	–	0	1	1	–
–	–	1	–	1	–
–	–	1	–	–	1
–	–	–	–	0	1
–	1	1	0	–	–
–	1	–	0	0	–
–	–	–	1	–	–
–	–	–	–	0	–
–	–	–	–	–	1
1	0	1	–	0	–

F	E	D	C	B	A
1	–	0	1	–	–
–	–	0	1	1	–
–	–	1	–	1	–
–	–	1	–	–	1
–	–	–	–	0	1
–	1	–	0	0	–
–	–	1	–	–	1
–	–	1	1	1	–
–	1	1	0	0	–
–	–	1	–	1	1
–	1	1	0	–	1

F	E	D	C	B	A	
1	–	0	1	–	–	$(C+\bar{D}+F)$
–	–	–	1	1	–	$(B+C)$
–	–	–	–	0	1	$(A+\bar{B})$
–	1	–	0	0	–	$(\bar{B}+\bar{C}+E)$
–	–	1	–	–	1	$(A+D)$

$$X=(C+\bar{D}+F)(B+C)(A+\bar{B})(\bar{B}+\bar{C}+E)(A+D)$$

Example 2: (The digits representiug the variables will be reversed to those in the previous cases. This has been done deliberately to illustrate that the variables may be initially assigned to digits in any order. By assigning the corresponding variables to the correct digits throughout the procedure the order has no effect on the conversion.):

$$X=[AB\bar{C}(D+E+F)(C+\bar{G}+H)]D\bar{G}+[AGH+(\bar{A}+\bar{B}+F)(\bar{C}+\bar{D}+GH)]$$

TO AND FORM

(1)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
1	–	–	–	–	–	1	1	P
0	0	–	–	–	0	–	–	S
–	–	0	–	–	–	–	–	P
–	–	–	0	–	–	–	–	P
–	–	–	–	–	–	1	1	P

(2)

A	B	C	D	E	F	G	H	
1	1	0	1	–	–	0	–	
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
1	–	–	–	–	–	1	1	
0	–	0	–	–	–	–	–	
0	–	–	0	–	–	–	–	
0	–	–	–	–	–	1	1	
–	0	0	–	–	–	–	–	
–	0	–	–	–	–	1	1	
–	–	0	–	–	0	–	–	
–	–	–	–	0	–	0	–	
–	–	–	–	–	0	1	1	

(3)

A	B	C	D	E	F	G	H	
1	1	0	1	–	–	0	–	
–	–	1	–	–	–	0	1	S
–	–	–	–	–	–	1	1	
0	–	0	–	–	–	–	–	
0	–	–	0	–	–	–	–	
–	0	0	–	–	–	–	–	
–	0	–	0	–	–	–	–	
–	–	0	–	–	0	–	–	
–	–	–	0	–	0	–	–	

(4)

A	B	C	D	E	F	G	H
1	1	0	1	–	–	0	–
–	–	–	–	–	–	1	1
0	–	0	–	–	–	–	–
0	–	–	0	–	–	–	–
–	0	0	–	–	–	–	–
–	0	–	0	–	–	–	–
–	–	0	–	–	0	–	–
–	–	–	0	–	0	–	–

(5)

A	B	C	D	E	F	G	H	
–	–	0	1	–	–	0	–	$\bar{C}D\bar{G}$
–	–	–	–	–	–	1	1	GH
0	–	0	–	–	–	–	–	$A\bar{C}$
0	–	–	0	–	–	–	–	$A\bar{D}$
–	0	0	–	–	–	–	–	$B\bar{C}$
–	0	–	0	–	–	–	–	$B\bar{D}$
–	–	0	–	–	0	–	–	$\bar{C}F$
–	–	–	0	–	0	–	–	$\bar{D}F$

$$X=\bar{C}D\bar{G}+GH+A\bar{C}+A\bar{D}+B\bar{C}+B\bar{D}+\bar{C}F+\bar{D}F$$

TO OR FORM

(1)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
1	–	–	–	–	–	1	1	P
0	0	–	–	–	0	–	–	S
–	–	0	–	–	–	–	–	S
–	–	–	0	–	–	–	–	S
–	–	–	–	–	–	1	1	P

(2)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
1	–	–	–	–	–	1	1	P
0	0	–	–	–	0	–	–	S
–	–	0	0	–	–	1	–	
–	–	0	0	–	–	–	1	

(3)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
–	–	0	0	–	–	1	–	
–	–	0	0	–	–	–	1	
0	0	–	–	–	0	1	–	
0	0	–	–	–	0	–	1	

(4)

A	B	C	D	E	F	G	H	
–	–	0	0	–	–	1	–	
–	–	0	0	–	–	–	1	
0	0	0	–	–	–	0	1	
0	0	0	–	–	0	–	1	
0	0	–	1	1	0	1	–	
0	0	–	1	1	0	–	1	
0	0	1	–	–	0	0	1	
0	0	–	1	–	0	1	–	
0	0	–	1	–	0	–	1	
0	0	–	–	–	0	0	1	

(5)

A	B	C	D	E	F	G	H	
–	–	0	0	–	–	1	–	$(\bar{C}+\bar{D}+G)$
–	–	0	0	–	–	–	1	$(\bar{C}+\bar{D}+H)$
0	0	–	1	–	0	1	–	$(A+B+D+F+G)$
0	0	–	–	–	0	0	1	$(A+B+F+\bar{G}+H)$

$$X=(\bar{C}+\bar{D}+G)(\bar{C}+\bar{D}+H)(A+B+D+F+G)(A+B+F+\bar{G}+H)$$

References

1. The Map Method for Synthesis of Combinational Logic Circuits. **M. Karnaugh.** *AIEE Transactions*, vol. 72 (*Communication and Electronics*), pt. I, Nov. 1953, pp. 593–99.

2. The Problem of Simplifying Truth-Functions. **W. V. Quine.** *The American Mathematical Monthly*, Menasha, Wis., vol. 59, Oct. 1952.

3. Minimization of Boolean Functions. **E. J. McCluskey, Jr.** *Bell System Technical Journal*, New York, N. Y., vol. 35, Nov. 1956, pp. 1417–44.

Discussion

Peter L. Randlev (IBM Research, Yorktown Heights, N. Y.): While Mr. Howard has been correct in attempting to find a simple straightforward method to obtain the most economical switching circuit, his method does not consistently do this.

The most economical form of a combinational circuit is dependent upon the components to be utilized. However, in developing a method for determining the configuration of such a circuit, it is not sufficient just to specify whether an alternational (sum of products) or a conjunctive (product of sums) form of logic shall be used. Often, the most economical configuration should be a combination of both these forms. It is also important to make provision for additional rules to obtain this configuration in view of: 1. the cost of each term in an expression as well as the cost of each literal, 2. the relative cost of terms composed of various quantities of literals when, as is most often the case, this cost is not linear with the quantity of literals, and 3. the cost of a circuit configuration composed of several levels of logic as would be obtained if the function were in a factored rather than a normal (pure) form.

Any method that is to be able to find the most economical circuit configuration must be able to take into account factors such as these for defining a minimum. The author has no definition of a minimum.

For any method to be easily and successfully programmed for a digital computer application, as Mr. Howard suggests, it must be both systematic and unambiguous. As presented, his method fails this criterion. The rules are not stated explicitly enough and, rather than being rigidly systematic, the examples themselves are demonstrations that the rules allow any of several sequences in performing the operations. To be used with a digital computer, it would obviously be necessary to use some means other than lines of several lengths to indicate the sequence of groupings required in transformation. This system does not even appear to be the best with hand methods as the comparison of the relative lengths of the connecting lines would become quite difficult to see in large problems. However, the author should be able to modify quite simply his rules and notation for use with a digital computer.

Other than the purely clerical operations of recording the original function and writing the final expression, the method may be divided into three general operations: 1. transformation or combination, 2. simplification, and 3. minimization.

The transformation operation changes the original function into a normal (pure) form. As desired, this would be either the alternational normal form or the conjunctive normal form. The simplification operation apparently is intended to find the terms that have the fewest possible literals. This is accomplished by pairing the given terms and removing those terms that are found to be completely contained within one other term. However, the rules do not always allow the term with the fewest literals to be found. The following is such an example:

Example A:

Original data	Desired answer
– 0 1 0	– 0 1 0
– 1 0 1	– 1 0 1
1 – 1 1	1 – – –
1 – 0 0	
1 1 1 –	
1 0 0 –	

The stated rules will not allow any simplification of this original function.

It is probably Mr. Howard's intention to define minimization as that expression of the function whose sum of the total quantity of literals in all terms plus the quantity of terms is the smallest possible. As the author has apparently recognized (in the last paragraph under "Obtaining the Minimum Form of a Solution"), his method does not necessarily find a minimum solution under this definition. The operation the author has described for use in minimization is to remove redundant terms; that is, terms that are expressed by two other terms. This method often removes a term that should appear in the final answer. For example, see the following:

Example B:

Original data	Desired answer
1 0 1 – –	1 0 1 – –
– 1 1 0 1	– 1 1 0 1
1 1 – 0 0	1 1 – 0 0
0 – 1 1 0	0 – 1 1 0
0 0 – 1 1	0 0 – 1 1
1 – 1 0 –	
– 0 1 1 –	

If the rules are followed consistently, the first term of the original data will be removed by the last two terms which will themselves remain intact. This nonminimal answer would be

– 1 1 0 1
1 1 – 0 0
0 – 1 1 0
0 0 – 1 1
1 – 1 0 –
– 0 1 1 –

In other cases these same rules fail to remove redundant terms, as in the following

Example C:

Original data	Desired answer
1 – 1 1	1 – 1 1
1 – 0 0	1 – 0 0
– 1 1 0	– 1 1 0
– 1 0 1	– 1 0 1
1 1 – –	

When the rules are more precisely stated as to the sequence of operations for use with a digital computer, the sequence of the original data becomes important in attempting to find a minimum solution. This is demonstrated by example *D*. If this problem is "minimized" with the data both in its original and in an inverted sequence, two answers of different cost are obtained.

Example D:

Original Data	Desired answer	Nonminimal "minimized" answer
1 1 – 1	1 1 – 1	1 0 1 –
1 0 1 –	1 0 1 –	0 1 0 –
0 1 0 –	0 1 0 –	1 – 1 1
1 – 1 1		– 1 0 1
– 1 0 1		

In the reference 2 cited by the author, Professor Quine discusses the minimization of problems he calls "irredundant." Professor Quine's example is

Example E:

Original data	Desired answers
1 0 –	1 0 – 0 1 –
0 1 –	0 – 1 or 1 – 0
– 1 0	– 1 0 – 0 1
– 0 1	

As Mr. Howard's method is completely dependent on finding redundant terms, it cannot find the minimum solution from these irredundant functions. The desired answers of these five examples can all be checked by other means (such as a truth table) to insure that they are equivalent to the original data. These problems are only samples of those problems that cannot be minimized by the author's rules.

It is my experience that as the problems become larger and more complex, finding a problem belonging to a class which cannot be minimized by this method becomes a more frequent occurrence. For a real minimization process to be useful, it must be possible to recognize the nonminimal cases even if a minimum answer cannot always be found (so that they may be solved by other methods).

There is no step in the author's method in which the intermediate results become more complex in form than the original. Thus, rather than referring to this as a minimization process, it would be more acceptable to call this portion of the author's paper a method to remove redundancy.

Extensions of the method to allow it first, to recognize cases where it cannot find a minimum, or second, to find a minimum answer for all cases, will make the process more complex and lengthy rather than reduce the necessary procedure. It is also necessary that a minimization procedure include the ability to handle "don't-care" conditions. These are normally created by specific input combinations that can't exist and taking them into consideration often allows drastic reductions in the required circuitry.

This method becomes, in common with all other methods known to this discusser, extremely difficult to use by hand for problems much larger than those that can be simplified by inspection, just because the problems are large. However, the large-problem area is the one in which the greatest savings may be made by the use of simplification schemes and which, therefore, requires the most work.

J. W. Stuart (Industrial Nucleonics Corporation, Columbus, Ohio): Mr. Howard has done a thorough job in presenting a method of converting and simplifying combinational logic circuits. Circuit design based upon the use of logic functions has received increasing attention in recent years.

The binary nature of the normal industrial control provides a comparatively simple logical system. It can be shown that the laws of "symbolic logic" are applicable in a rather restricted sense. Specifically, devices capable of providing the logical properties of assertion and negation are the basic requirement. Assertion corresponds with ON while negation is representative of OFF.

Assertion can be categorized into two principal groups. Alternative assertion is described by the verbalization OR whereas co-ordinate assertion (coincidence) is descriptive of AND. Negation is by definition an overt representation of the absence of assertion. This inversion is often termed the complementary function. The practical importance of grasping the concept of negation cannot be overemphasized. Negation is not a passive zero state. For example, a normally closed contact does not provide negation unless it is also connected to a voltage source. Furthermore an unoperated normally open contact is not synonymous with negation, only compatible with it. An active signal must be provided to represent the open condition of this normally open contact in its unoperated state. Identical reasoning applies to static switching devices.

It is possible and practical to construct entire logical systems, including memory, utilizing only negation and one of the two assertions. Therefore, a given system might contain only OR and NOT. An alternative system could contain only AND and NOT. It should be evident that a system lacking negation, or NOT, is not possible.

A system composed of all three basic logic functions is defined as an English logic type. It provides the most flexible system design choice and is apparently the closest approximation of normal thought as evidenced by the language structure of at least the English-speaking people. From a practical standpoint, such a logical system is easier to understand, design, and troubleshoot. This consideration is important since the basic description of the operation sequence is invariably presented in a verbalized form. Furthermore, the usual control design methods depend on this flexibility unless a mathematical tool such as the algebra of logic (Boolean algebra) is utilized. Thorough mastery of this tool is necessary to allow effective design. This algebra is limited in over-all application since time variant or sequential operation is difficult to take into account. Therefore, the entire class of memory types is excluded. However, for optimizing extremely large systems or very complex strings of combinational logic configurations, this algebra is very effective and worth while.

The usual starting place in tackling a control design is with a formulation of the conditions necessary to activate each output. Upon completion of the individual circuits as well as after final integration of the over-all system, it is desirable to expend a reasonable effort in checking and simplification. The art of simplification is tied closely to the economics of alternative logic components available. In the case of the static-switching system under consideration, the most economical system is generally achieved by the maximum use of OR units. Furthermore, the manner in which memory units are formed has an important bearing upon formation of optimum circuit design. Broadly, two processes can be used to affect simplification of combination circuits composed of ORS, ANDS, and NOTS or variations thereof. These processes are conversion and inversion. Conversion is concerned principally with OR–AND circuits, while inversion is concerned with NOT functions as the term implies.

A given portion of a combinational circuit may be the termination of two or more ANDS at a single OR. Conversely, it may be the termination of two or more ORS at a single AND. It can be shown that a one-to-one correspondence exists between these alternative circuits such that one may be converted to the other with about an even chance that simplification will result.

The greatest possibility for worth-while simplification exists whenever a given combinational circuit contains a repetition of one or more identical signals. The final circuit may or may not contain repetitive signals since economy is not primarily a logical consideration. It is equally important to note that if a signal A and its negation or complement $\bar{A}$ when found in a given combination are often but not invariably redundant.

Frequently, the availability of the normally closed contacts of sensing devices permits simplification by permitting inversion of the circuit. Inversion entails changing OR conditions to ANDS and AND conditions to ORS with all input signals inverted (negated) and providing the complement, or NOT, of the resultant circuit.

Both conversion and inversion may be carried out on a given circuit for the ultimate in simplicity. The type of memory circuit with which the combinational circuit is connected in a large measure determines whether conversion, inversion or both are profitable.

This paper presents a complete and usable method of handling OR–AND conversions and performing a straightforward procedure of simplification.

Memory circuits would appear to be most effectively handled with sequence or time-bar charts to define the combinational functions which occur momentarily that must be remembered for subsequent use in the circuit operation. Unless such a method is used to reduce the theoretical memory possibilities to actual requirements, the task is almost unmanageable. At any rate I have not seen a really practical method of handling memory functions with the use of Boolean algebra alone. The analysis of memory situations certainly should present a fruitful field for future investigation.

R. L. Howard: The author wishes to express his gratitude to Mr. Stuart and Mr. Randlev for their discussions.

The author's background is basically that of industrial-control systems while Mr. Randlev's is predominantly concerned with digital computer applications. His discussion therefore provides an appraisal of the procedure from the computer viewpoint, which is an area with which the author is least acquainted. Primarily Mr. Randlev's comments are directed to certain generalities contained in the paper that would seem to imply that the scope of the procedure presented is unlimited. It was not the author's intention to create this false impression, particularly regarding the minimization of Boolean functions.

The procedure presented does not provide methods for obtaining the most economic solution for a particular application or function. Surely many factors must be taken into consideration before the optimum form can be determined and it has been assumed that the designer has solved this portion of the problem before utilizing the procedure. Those applications that do find their most economic solution in one of the pure logic forms can be resolved directly from the procedure. Other applications that find their optimum solution in the combinational form, will require additional effort on the part of the designer in using the procedure as is the case with all similar methods known to the author.

Different lengths of lines are utilized throughout the procedure to indicate groupings within the function as well as serving as operational devices. This does not mean that the procedure is restricted to these symbols any more than the AND or OR functions in Boolean algebra are always represented by the same symbol. The author readily agrees that other designations can be utilized for the desired indication and that these may allow easier handling of the procedure.

The "don't-care" type of term has not been included in the procedure and what would be involved in taking this into consideration can only be determined by further investigation. In the same vein, it is the author's contention that only further examination can determine the length of the procedure necessary to enable it to derive an absolute minimum for a particular function.

It is a well-known fact that vast amounts of study have been directed toward the area of minimizing Boolean functions. Only a fraction of this effort could be covered in a paper of this length. Being aware of this, I made an attempt to qualify the rules for minimization, but obviously this qualification was not stressed to a great enough degree. Mr. Randlev's comments on the simplification and minimization portion of the paper have pinpointed the restriction of the rules covering these sections and have brought these limitations firmly to the attention of the author and the reader. These comments (including the author's presumed definition of minimization) are accurate and are invaluable to the author and the paper.

In summarizing Mr. Randlev's discussion, I do not feel that the rules presented in the paper have been shown to be incorrect but rather that they do not include all cases.

Mr. Randlev's and Mr. Stuart's discussions have been of great assistance. It is constructive analysis such as this that will increase the value of the procedure as a tool in logic design.

A Stability Criterion for Nonlinear Systems

Y. H. KU
FELLOW AIEE

A. A. WOLF
MEMBER AIEE

IN ANOTHER PAPER[1] Wolf reports on the recurrence relations in the solution of a certain class of nonlinear systems. In Wolf's dissertation[2] a mathematical theory for the analysis of a class of nonlinear systems has been developed. This paper reports on a stability criterion based on the aforementioned dissertation and its application to nonlinear as well as linear feedback control systems.

The method of stability analysis consists essentially of first finding the response of a nonlinear system to a given forcing function by either the method of partition given in reference 1 or the Taylor-Cauchy transform method.[3] These methods give the solution of certain class of nonlinear systems[1,2] in the form of a power series, such that the response $x(t)$ is governed by a function of the exponential type given in the following equation.

$$x(t) = \sum_{n=0}^{\infty} h_n t^n \tag{1}$$

where h_n are constants for different values of n.

The Laplace transform of $x(t)$ is

$$X(s) = \sum_{n=0}^{\infty} \frac{h_n n!}{s^{n+1}} \tag{2}$$

where s is a complex variable.

Stability Criterion

The stability criterion is given as follows. Given $x(t)$ the solution of a nonlinear differential equation (which represents a given nonlinear physical system) such that $x(t) = 0$ (ϵ^{st}). Then $x(t) \to 0$ as $t \to \infty$ if and only if the convex singularity hull (CSH) of $X(s)$ is entirely in the left-half plane (LHP).

The CSH is constructed by joining the points of singularity of the transform $X(s)$ in the s-plane. It may be noted that the singularities of $X(s)$ are not necessarily poles.

The nonlinear differential equation can be given in the following form

$$Z(D)x(t) + F(x, x', \ldots) = r(t) \tag{3}$$

where

$Z(D)$ = linear differential operator
$F(x, x', \ldots)$ = nonlinear function of x, its derivative x', ...
$r(t)$ = the forcing function or input
$x(t) = c(t)$ = the response
$D = d/dt$ = the differential operator
$D^2 = d^2/dt^2$, ...
$x' = dx/dt$ = the derivative of the response with respect to time t

A nonlinear system is said to be stable if to every bounded-decaying driving function or input and for all initial conditions the response $x(t)$ approaches zero as time increases to infinity. If the foregoing condition is not satisfied, the system may be oscillating in the steady-state or may run away and become unstable.

In a linear system no mention is usually made of bounded values when speaking of the stability of the system. The reason for this is that a linear system has a response whose form is essentially independent of the initial conditions. In a nonlinear system, it may go from a stable to an unstable state during some transition, such as switching which may alter the initial conditions. In such cases the nonlinear system is considered conditionally stable and the range of input functions and initial conditions must be specified for stable operation.

In linear systems the CSH is composed of poles for lumped parameter systems. For stability these poles must be confined to the LHP. In nonlinear systems the CSH is composed of singularities (which are not necessarily poles) and for stability these singularities must be confined to the LHP. Thus the stability criterion given previously holds for nonlinear as well as linear systems.

Construction of the CSH

The CSH and other related concepts in this paper are obtained from entire function theory.[4] For the purpose of this paper, however, the following simple examples show how CSH can be obtained for stability studies of physical systems.

Example 1

$$X(s) = \sum_{n=0}^{\infty} \frac{1}{s^{n+1}} \tag{4}$$

This is a special form of equation 2, with $h_n n! = 1$. Equation 4 or 2 (in general) looks like a Laurent series. However, one must not conclude that at $s = 0$, $X(s)$ has an essential singularity. Instead it may be that $X(s)$ at $s = 0$ is a regular point.

Clearly equation 4 is an expansion of

$$X(s) = \frac{1}{s-1} \tag{5}$$

At $s = 0$, $X(s) = -1$, while at $s = 1$, $X(s)$ has a pole. So the series of equation 4 converges outside of the circle of convergence whose radius is $R = 1$. In general, if the power series of equation 2 converges outside of the circle of convergence whose radius is $R = a$, one must not conclude that at $s = 0$, $X(s)$ has an essential singularity.

Since the singularity $s = 1$ is not in LHP the system is unstable. This can be verified by taking the inverse transform of equation 5.

Example 2

$$X(s) = \sum_{n=0}^{\infty} \frac{a^n}{s^{n+1}} \tag{6}$$

Compared with equation 2, $h_n n! = a^n$, where a is a constant. Equation 6 is an expansion of

$$X(s) = \frac{1}{s-a} \tag{7}$$

At $s = a$, $X(s)$ has a singularity, which is a pole. The system is unstable.

Example 3

$$X(s) = \sum_{n=0}^{\infty} \frac{(-1)^n a^n}{s^{n+1}} \tag{8}$$

is an expansion of

$$X(s) = \frac{1}{s+a} \tag{9}$$

At $s = -a$, $X(s)$ has a pole, which is in LHP. Hence the system is stable.

Example 4

Let the simplified from of a given power series be given by

Paper 59-231, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 15, 1958; made available for printing December 15, 1958.

Y. H. KU is with the Moore School of Engineering, University of Pennsylvania, Philadelphia, Pa.; A. A. WOLF, formerly with the Moore School of Engineering, is now with the Stromberg Carlson Company, Division of General Dynamics, Rochester, N. Y.

The authors wish to acknowledge the support given by the National Science Foundation for their research on nonlinear analysis.

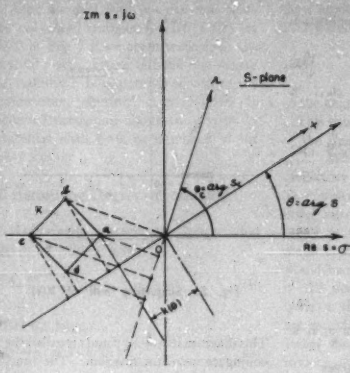


Fig. 1 (left). CSH for a stable system

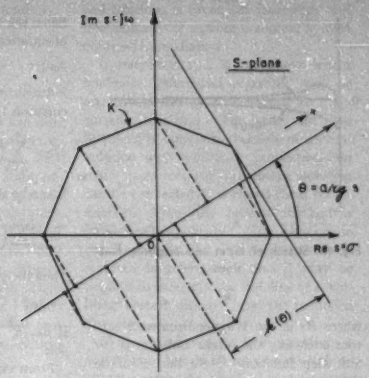


Fig. 2 (right). CSH for an unstable system. Set of points K and its projection on $\theta = \arg s$

$$X(s) = \frac{1}{s+2} + \frac{1}{s+1-j2} + \frac{1}{s+1+j2} \tag{10}$$

The singularities are at $s=-2$, $s=-1+j2$, and $s=-1-j2$. Connecting all three points gives the CSH in the form of a triangle. Since the CSH lies entirely in the LHP, the system is stable.

The foregoing simple examples are taken from the linear systems so that the construction of the CSH can be demonstrated in a straightforward manner. Fig. 1 shows a CSH for a stable system, while Fig. 2 shows a CSH for an unstable system.

Radius of Convergence and Direction of Singularity

From equations 4, 6, and 8 in examples 1, 2, and 3 of the foregoing, it is seen that the origin in the s-plane is not included as a singularity. In example 1, the circle of convergence has a radius $R=1$. In examples 2 and 3, the circle of convergence has a radius $R=a$. It is noted that in examples 1, 2, and 3 the singularity lies on the circle of convergence. In general, the radius of the circle of convergence is given by

$$R = \overline{\lim_{n\to\infty}} \left| n!\, h_n \right|^{1/n} \tag{11}$$

In example 1, equation 11 gives $R=1$. In examples 2 and 3 equation 11 gives

$$R = \overline{\lim_{n\to\infty}} \left| n!\, h_n \right|^{1/n} = \overline{\lim_{n\to\infty}} \left| a^n \right|^{1/n} = a \tag{12}$$

The direction of the singularity (which lies on the circle of convergence) is denoted by an angle θ, which can be obtained as follows. Let

$$d_k = \sum_{n=0}^{k} \binom{k}{n} e^{jn\theta}\, \beta_{n+1} \tag{13}$$

where

$$\beta_{n+1} = \frac{n!\, h_n}{R^{n+1}}$$

and

$$\binom{k}{n} = \frac{k!}{n!(k-n)!} \tag{14}$$

Then θ is the direction of the singularity with radius R if

$$\overline{\lim_{k\to\infty}} \left| d_k \right|^{1/k} \geq 2 \tag{15}$$

With the radius of the circle of convergence obtained from equation 11 and with the direction of singularity determined by equations 13 and 15, the singular point lies at

$$s = Re^{j\theta} \tag{16}$$

The singularity given by equation 16 is then subtracted from $X(s)$ given by equation 2 and the process is repeated until all singularities are removed.

From example 1, $\beta_{n+1}=1$, and

$$d_k = \sum_{n=0}^{k} \binom{k}{n} e^{jn\theta} \tag{17}$$

Then

$$\overline{\lim_{k\to\infty}} \left| \sum_{n=0}^{k} \binom{k}{n} e^{jn\theta} \right|^{1/k} = \overline{\lim_{k\to\infty}} \left| (1+e^{j\theta}) \right|^{1/k} = \left| 1+e^{j\theta} \right| \tag{18}$$

It is seen that the right-hand member of equation 18 is equal to 2 for $\theta=0$.

Application to a Nonlinear Feedback System

Consider the nonlinear feedback system of Fig. 3. The differential equation governing the dynamics of the system is given in the following

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + x + x^2 = r(t) \tag{19}$$

The forcing function or input as shown Fig. 4 and given by

$$r(t) = e^{-4t} - 2e^{-3t} \tag{20}$$

Assume that the initial conditions are: $x(0)=0$ and $x'(0)=1$. Substituting equation 20 in equation 19 and partitioning at the first linear term gives

$$D^2x = \frac{d^2x}{dt^2} = e^{-4t} - 2e^{-3t} - (2x' + x + x^2) \tag{21}$$

The auxiliary equation is[1,2]

$$\sum_{n=0}^{\infty} C_n t^n = e^{-4t} - 2e^{-3t} - (2x' + x + x^2) \tag{22}$$

and the solution of equation 21 is given by[1,2]

$$x(t) = \sum_{n=0}^{\infty} C_n Q_n(t) + K_0 + K_1 t \tag{23}$$

where

$$Q_n(t) = \int_0^t \tau^n y(t-\tau)\, d\tau \tag{24}$$

and

$$y(t) = \frac{1}{2\pi j} \int_{B_T} \frac{e^{st}}{s^2}\, ds = tu(t) \tag{25}$$

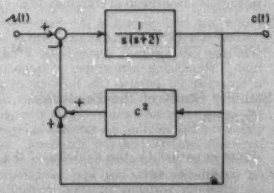


Fig. 3. Nonlinear feedback system

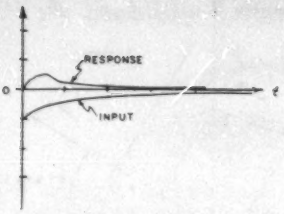


Fig. 4. Sketch of input and response functions

where Br is the Wagner-Bromwich contour enclosing the origin and $u(t)$ is the unit step function. Note that s^2 inside the integral is a special form of $Z(s)$, which depends upon the form of $Z(D)$ in equation 3. In equation 21, $Z(D)=D^2$; hence $Z(s)=s^2$. Substituting equation 25 in equation 24 gives

$$Q_n(t)=\int_0^t \tau^n(t-\tau)d\tau=\frac{t^{n+2}}{(n+1)(n+2)} \qquad (26)$$

Applying the initial conditions gives $K_0=0$ and $K_1=1$ in equation 23. Thus

$$x(t)=\sum_{n=0}^{\infty} C_n\frac{t^{n+2}}{(n+1)(n+2)}+t \qquad (27)$$

Substituting eqution 27 in equation 22 and solving for the constants recursively[1] gives

$$C_0=-3,\ C_1=7,\ C_2=-15/2,\ C_3=31/6,\ \ldots \qquad (28)$$

Substituting these values in equation 27 gives

$$x(t)=t-\frac{3}{2}t^2+\frac{7}{6}t^3-\frac{15}{24}t^4+\frac{31}{120}t^5-\ldots \qquad (29)$$

The n th term is recognized as

$$\frac{(-1)^{n+1}(2^n-1)}{n!}t^n$$

so that

$$x(t)=\sum_{n=0}^{\infty}\frac{(-1)^{n+1}(2-1)}{n!}t^n \qquad (30)$$

The Laplace transform of equation 30 is

$$X(s)=\sum_{n=0}^{\infty}\frac{(-1)^{n+1}(2^n-1)}{s^{n+1}} \qquad (31)$$

Stability Study of the Feedback System

From equation 11, the modulus of the first singularity is

$$R_1=\lim_{n\to\infty}\left|(-1)^{n+1}(2^n-1)\right|^{1/n}=2 \qquad (32)$$

Since the radius is not unity, application of equation 14 gives

$$\beta_{n+1}=-(-1)^n(2^n-1)2^{-n-1} \qquad (33)$$

Substituting equation 33 for β_{n+1} in equation 13 gives

$$d_k=-\sum_{n=0}^{k}\binom{k}{n}(-1)^n(2^n-1)2^{-n-1}e^{jn\theta} \qquad (34)$$

It can be shown that for $\theta=\pi$

$$\lim_{k\to\infty}\left|\sum_{n=0}^{k}\binom{k}{n}2^{-n-1}(2^n-1)\right|^{1/k}=2 \qquad (35)$$

Thus the singularity is at

$$s_1=2e^{j\pi} \qquad (36)$$

If $\beta_{n+1}R^{n+1}$ is a power function s_1 is a pole.

From example 3, equation 8 gives

$$X_1(s)=\frac{1}{s+2}=\sum_{n=0}^{\infty}\frac{(-1)^n2^n}{s^{n+1}} \qquad (37)$$

for $a=2$. Subtracting $X_1(s)$ from $X(s)$ as given by equation 31 yields

$$X_2(s)=X(s)-X_1(s)=-\sum_{n=0}^{\infty}\frac{(-1)^n}{s^{n+1}} \qquad (38)$$

The singularity of equation 38 is at

$$s_2=e^{j\pi} \qquad (39)$$

by inspection. Plotting the singularity hull, shown in Fig. 5, and applying the stability criterion, the system is stable.

In this example, the singularities are poles. In simplified form, equation 31 corresponds to

$$X(s)=\frac{1}{s+1}-\frac{1}{s+2} \qquad (40)$$

Taking the inverse transform of equation 40 gives

$$x(t)=e^{-t}-e^{-2t} \qquad (41)$$

which is the solution of equation 21 in closed form. The first derivative is

$$x'(t)=-e^{-t}+2e^{-2t} \qquad (42)$$

At $t=0$, equations 41 and 42 give respectively $x(0)=0$ and $x'(0)=1$.

Conjugate Indicator Function and Diagram

So far the tranform pairs $x(t)$ and $X(s)$ have been considered. Suppose that t be a complex variable denoted by

$$t=re^{j\theta} \qquad (43)$$

where $r=|t|$. Consider Fig. 6. The growth of $x(t)$ is determined by the indicator function

$$f(\theta)=\overline{\lim_{r\to\infty}}\ \frac{\log|x(t)|}{t} \qquad (44)$$

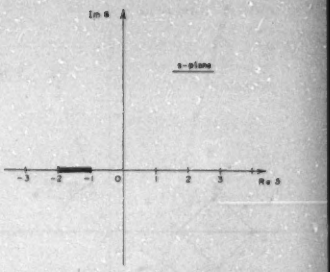


Fig. 5. Singularity hull of X(s)

This diagram involving θ and t is called the conjugate indicator diagram. The function $f(\theta)$, which is also called the Phragmén-Lindelöf function, determines the different behaviors of $x(t)$ in the different directions of t as $r\to\infty$. To understand the meaning of $f(\theta)$ consider

$$x(t)=e^{(a+jb)t} \qquad (45)$$

Substituting equation 43 in equation 45 and simplifying gives

$$|x(t)|=e^{r(a\cos\theta-b\sin\theta)} \qquad (46)$$

The behavior of $|x(t)|$ depends first of all on $|t|$, which is independent of θ. Then as $t\to\infty$ the growth of $x(t)$ depends on the sign of the factor involving θ. That is to say, along different rays of t (as defined by equation 43), the value of θ determines whether $(a\cos\theta-b\sin\theta)$ is negative or positive. If it is negative it is clear that $|x(t)|\to 0$ as $r\to\infty$. It is also clear that there is a range of θ for which the factor $(a\cos\theta-b\sin\theta)$ is positive. In these cases $|x(t)|$ will be unbounded as r tends to infinity. The function $f(\theta)$ is thus a measure of whether $|x(t)|$ is stable or not. It may be noted here that the properties of $f(\theta)$ are the basis for the development of the stability criterion given in the foregoing.

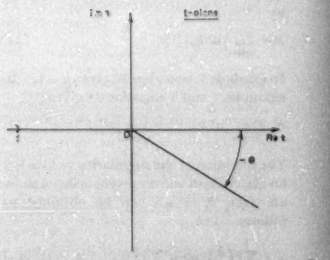


Fig. 6. Conjugate indicator diagram

Since we deal with positive real time, we only need consider results of $f(\theta)$ for $\theta=0$ in Fig. 6 if we are interested in the stability of a system but not the degree of stability. It is possible to utilize the information obtained from $f(\theta)\neq 0$ to obtain information about the degree of stability when $\theta=0$, as discussed in the next section.

Determining Degree of Stability

Referring to equation 46, $f(\theta)$ is roughly given as follows: Since

$$\frac{\log |x(t)|}{r} = a\cos\theta - b\sin\theta \tag{47}$$

then for large r:

$$f(\theta) \approx a\cos\theta - b\sin\theta \tag{48}$$

It is seen that if the system is stable then for $\theta=0$, $f(\theta)\approx a<0$. An angle $\theta=\theta_c\neq 0$ exists such that $f(\theta_c)>0$. We now show how to find this θ_c graphically from the CSH.

Suppose that the CSH is in the LHP, as shown in Fig. 1. The critical angle θ_c then is the direction of the ray which has a projection of a singularity at the origin. This angle θ_c measures the direction when first the system has an unbounded complex time response. To determine this angle first plot the CSH in the s-plane. Construct the perimeter of the CSH by connecting points of singularity with straight-line segments. Draw a straight-line tangent to the perimeter (as shown by b in Fig. 1) from the origin. Next construct a line or perpendicular to line b0. The inclination angle of line 0r is θ_c.

Repeat this procedure for the tangent point in the lower-half plane $\omega<0$. The angle so obtained is the negative critical angle, to be denoted by $\theta_c{}^1$. The sum of these angles is the total critical angle θ_T. This leads to the definition

$$\eta = \frac{\theta_T}{\pi} \tag{49}$$

as the degree of stability.

This equation gives a first-order measure of the degree of stability. Since $\theta_T\leq\pi$, $\eta\leq 1$. When $\eta=1$ the system is nonoscillatory and exponentially damped. To compare two such systems a second-order measure of the degree of stability might be the magnitude of the displacement of CSH along the σ axis. This would be the magnitude of the distance of the nearest singularity to the origin lying on the real axis σ. A third measure would be the distribution of the projected singular points along the critical ray.

Conclusions

In this paper we have presented some aspects of a stability theory and a generalized stability criterion for the analysis of a certain class of nonlinear systems.

In nonlinear systems the stability is markedly dependent upon the boundary values (the initial conditions, for example). It is also dependent upon the magnitude of the driving function. In general, the range of amplitudes of the driving function and the range of initial or boundary conditions would have to be specified. A stability criterion for nonlinear systems must involve all these conditions in addition to the characteristics of the system. The proposed criterion based upon the concept of the convex singularity hull contains all these conditions automatically.

Because of the conformal properties of the CSH and the conjugate indicator diagram, these diagrams can be superimposed and relations between the frequency and time domains of a given nonlinear system obtained immediately.

In references 1, 2, 3 it is shown how a nonlinear differential equation such as equation 3 can be transformed to a recurrence equation. The latter can then be solved in open form because of its convolution and symmetry properties. In many instances the recurrence solution can be converted to a closed form by induction. Then the present method is immediately applicable. It appears from the Taylor-Cauchy inversion transform reported in reference 3 that it may be possible to ascertain the stability of a nonlinear system directly from the recursion formulas.

If the closed form of h_n in equations 1 and 2 is not known, it may be possible to obtain roughly the stability of the system by taking only the first few terms of the expansion of $x(t)$. In this case we must conceive a new function which majorizes $x(t)$ at least in the known truncated portion. However, caution must be taken to avoid serious errors if the unknown or discarded part is not majorized by this new function.

References

1. Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems. Alfred A. Wolf. *AIEE Transactions*, vol. 78, pt. I (*Communication and Electronics*), 1959 (Paper no. 59-919).

2. A Mathematical Theory for the Analysis of a Class of Nonlinear Systems. A. A. Wolf. *Ph.D. Dissertation*. University of Pennsylvania, Philadelphia, Pa., June 1958.

3. Taylor-Cauchy Transforms for Solving Certain Class of Nonlinear Systems. Y. H. Ku, A. A. Wolf. *Convention Record*. Institute of Radio Engineers, New York, N. Y., 1959.

4. Entire Functions (book), R. P. Boas. Academic Press, New York, N. Y., 1954.

5. Theorie und Undwendung der Laplace Transformation (book), G. Doetsch. Dover Publishers, New York, N. Y., 1943.

Discussion

N. F. Tsang (University of Mississippi University, Miss.): The authors are to be congratulated for dealing thoroughly with the thorny subject of stability of nonlinear systems. In fact the criterion is applicable to any function, not limited to those derived from nonlinear systems.

The concept of the CSH is introduced. As the LHP is not a concave region, it seems that a simpler statement "all the singularities must be in the LHP" should be sufficient for our purpose.

There will be considerable difficulty both in deriving the Laplace Transform from the nonlinear equations with their boundary conditions, and in determining the singular points from a Laplace Transform in inverse power series of s. Equation 41 from equations 19, 20 with $x(0)=0$, $x'(0)=1$ is not typical for nonlinear systems. It would be very enlightening if the authors will apply the method in detail (first obtain a power series, then its Laplace Transform term by term and the singularities) to equations 19, 20 with the boundary conditions, say $x(0)=-1.9$, $x'(0)=0.2$.

Y. H. Ku and **A. A. Wolf**: The authors would like to thank Professor N. F. Tsang for his interesting comments on our paper, and take this opportunity to thank every one who discussed this paper with us at its presentation.

With regard to Professor Tsang's comment on the singularities, reasons shall be briefly given for the necessity of convexity of an entire function in different directions in the complex t-plane can be specified as indicated in reference 2 of the paper by the Phragmén-Lindelöf indicator function

$$fh(\theta) = \overline{\lim_{r\to\infty}} \frac{\log |x(re^{j\theta})|}{r^{\rho}} \tag{50}$$

where ρ denotes the order of the function and $t=re^{j\theta}$. It is known from the theory of entire functions that if $\rho\leq 1/2$ for $x(t)$ and if $x(t)$ is not a constant the minimum modulus is everywhere unbounded. That is to say that $x(t)$ denotes the response of an unstable system. In reference 2 of the paper it is shown that for certain class of physical systems ρ is unity or approximately so. Nonlinear physical systems discussed in this paper are governed by functions of exponential type and order one.

Putting $\rho=1$ in equation 50 gives the indicator function

$$fh(\theta) = \lim_{r\to\infty} \frac{\log |x(re^{j\theta})|}{r} \tag{51}$$

for this class of physical systems.

The indicator function is everywhere bounded by the type of $x(t)$, namely τ. The type τ of a function $x(t)$ is given by

$$\tau = \lim_{r\to\infty} \frac{\log M(r)}{r} \tag{52}$$

according to reference 2 of the paper, where $M(r)$ denotes the maximum modulus of $x(t)$ and $r = |t|$.

Hence

$$M(r) = \max_{r\in D_1} |x(t)| \tag{53}$$

where D_1 is the domain of convergence of $x(t)$ and r belongs to D_1. From equation 53 it is evident that

$$M(r) \geqslant |x(t)| \tag{54}$$

for all r belonging to D_1. Thus on substituting equation 54 into equation 52 and comparing the result with equation 51 one obtains the simple relation

$$fh(\theta) \leq \tau \tag{55}$$

Geometrically equation 55 signifies that $fh(\theta)$ is associated with a convex domain. Because of the conformal behavior of the complex t-plane and the complex s-plane, the distribution of singularities is also convex. Hence the concept of a CSH is important and useful.

This approach can be extended to those systems giving rise to nonconvex singularity hulls by suitably breaking the nonconvex sets into the sum of convex sets and then applying to each such convex set the theory presented in this paper.

In answering Professor Tsang's next comment, the example given in the paper shall be discussed with general boundary conditions:

$$x(0) = a = K_0;\ x'(0) = b = K_1 \tag{56}$$

where a and b are real constants. Equation 23 in the paper gives then

$$x(t) = \sum_{n=0}^{\infty} C_n \frac{t^{n+2}}{(n+1)(n+2)} + a + bt \tag{57}$$

The general recursion formula for C_n is given by

$$C_n = \frac{(-1)^n}{n!}[4^n - 2(3)^n] - \frac{2C_{n-1}}{n} - 2bC_n^{(0)} - \frac{C_{n-2}}{n(n-1)} - aC_n^{(0)} - bC_{n-1}^{(0)} - C_{n-4,\,0}^{(0)} - 2a\frac{C_{n-2}}{n(n-1)} - 2b\frac{C_{n-3}}{(n-1)(n-2)} - a^2C_n^{(0)} - 2abC_{n-1}^{(0)} - b^2C_{n-2}^{(0)} \tag{58}$$

where

$$C_m^{(0)} = \begin{cases} 1 & \text{when } m = 0 \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

$$C_{n-4,\,0}^{(2)} = \sum_{k=0}^{n-4} \frac{C_k C_{n-k-4}}{(k+1)(k+2)(n-k-3)(n-k-2)} \tag{60}$$

When $a=0$ and $b=1$, one obtains the recurrence relation (see equation 60) from which equation 28 in the paper is obtained.

$$C_n = \frac{(-1)^n}{n!}[4^n - 2(3)^n] - \frac{2C_{n-1}}{n} - 2C_n^{(0)} - \frac{C_{n-2}}{n(n-1)} - C_{n-1}^{(0)} - C_{n-4,\,0}^{(2)} - \frac{2C_{n-3}}{(n-1)(n-2)} - C_{n-2}^{(0)} \tag{61}$$

It is always possible to solve equations 59 and 60 recursively; i.e., for $n = 0, 1, 2, \ldots$. Suppose that this is done, then the nth term shall either be recognized or not. In the former case one proceeds as in the paper. In the latter case the form of $C_0, C_1, C_2, \ldots$ might be examined to see if it takes a particular form. One such form might be

$$C_n = \frac{A\alpha^n}{n!} \tag{62}$$

where A and α are constants, as suggested by the nth term given in the paper. If this should prove to be unfruitful, instead a majorizing or asymptotic solution for C_n might be obtained.

In conclusion, it is seen that it is a straightforward matter to obtain recurrence relations which can be solved. It is recognized that some difficulties may arise in getting the closed form of the nth term. Further efforts are being made in this direction. It may be remarked that the method of stability analysis can be reversed so that one can synthesize a nonlinear feedback control system with a prescribed frequency response. In this case the nth term is known as soon as the CSH is specified. It is even possible to build into the system a given degree of stability, a given settling time, and other relevant characteristics.

Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading

PRAPAT CHANDAKET
NONMEMBER AIEE

ALLEN B. ROSENSTEIN
MEMBER AIEE

MANY servo systems possess open-loop transfer functions which contain one or more pairs of complex conjugate poles. Often a pair of these complex conjugate poles will lie close to the imaginary axis and will cause the uncompensated system to possess a relatively poor damping ratio or damping exponent. Usually, satisfactory compensation of such systems can be achieved through the use of tachometer feedback or the conventional compensators possessing real axis singularities. However, there are occasions when complex conjugate singularity compensation would possess distinct advantages. This is particularly true if the compensating complex singularities can be introduced by a simple *R-C* (resistance-capacitance) passive network with zero d-c attenuation.

Evans[1] has shown how the simple bridged-T network can be used with a-c servos to produce a phase lead. In this article, the application of the elementary bridged-T network is developed for d-c system compensation. The locus of roots of the transfer function of the bridged-T network is presented in dimensionless form. It is found that the transfer function of the bridged-T possesses two zeros, one or the other of which may be adjusted to fall upon any desired point in the left half-plane including any complex conjugate pair. The bridged-T also produces a pair of poles which fall upon the negative real axis on either side of the real component of the pair of zeros.

A set of curves has been developed which allows very rapid synthesis of the unloaded bridged-T network with its compensating complex conjugate zeros falling at any desired left half plane

Paper 59-220, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 2, 1958; made available for printing December 16, 1958.

LIEUT.-COMDR. PRAPAT CHANDAKET is with the Royal Thai Navy, and ALLEN B. ROSENSTEIN is with the University of California, Los Angeles, Calif.

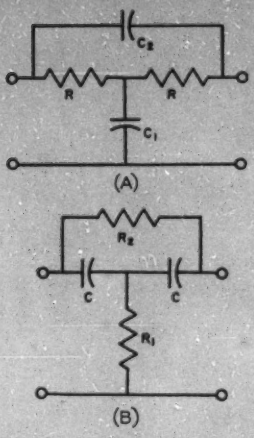


Fig. 1. Common types of bridged-T configuration

points. Both types of simple bridged-T network can be synthesized from the same graphs.

To investigate the effects of loading upon the bridged-T network, a general method has been developed to observe the locus of the transfer function poles and zeros of any four terminal passive network with varying load termination. The effects of three different types of bridged-T loading are investigated. It is found that these loads do not affect the location of the zeros but can be employed to change materially the location of the compensating poles upon the negative real axis.

The application of the loaded bridged-T network to a position servomechanism, such as a fin control servomechanism whose open-loop transfer function along with the airplane dynamics function, contains complex poles in the left half-plane, is found to offer material improvements in the over-all system performances.

Analysis of the Bridged-T Network

Transfer Function

Bridged-T *R-C* networks fall in two types of configurations; See Fig. 1. The network in (A) will be referred to as bridged-T type *Ia*, and the other as type *Ib*. Bridged-T type *Ia* has been analyzed by Evans[1] and others. Evans used root locus techniques to determine its pole-zero configuration and found it to be characterized by two poles and two zeros. These poles are always real and negative, but the zeros can be made negative real or complex, with negative real parts depending on the ratio of capacitances.

The root locus technique will be used to show the movement of the poles and zeros of the transfer functions of these networks. The bridged-T type *Ib* is selected for illustration. With zero input source impedance and infinite output load impedance, the voltage transfer function of the network type *Ib* is found to be

$$\frac{E_{out}}{E_{in}} = \frac{1+2R_1Cs+R_1R_2C^2s^2}{1+(2R_1+R_2)Cs+R_1R_2C^2s^2} \qquad (1)$$

Locus of Poles and Zeros

The equation for the zeros of equation 1 is given by

$$1+2R_1Cs+R_1R_2C^2s^2=0 \qquad (2)$$

which can be arranged as

$$\frac{1+2R_1Cs}{R_1R_2C^2s^2}=-1 \qquad (3)$$

The pole equation can also be arranged as

$$\frac{1+2R_1Cs}{R_2Cs(1+R_1Cs)}=-1 \qquad (4)$$

If $1/R_2$ is assumed to be the loop gain and allowed to vary from zero to infinity, the root locus plot of equation 3 will give the zero configuration for the network, while the root locus of equation 4 will yield the pole characteristics. These are shown in Fig. 2.

From Fig. 2, it may be seen that the zeros of the transfer function can be made complex or real in the left half-plane while the poles are always negative real when R_2 is varied.

Similarities

The transfer function of the network type *Ia* with zero input impedance and infinite output load impedance is found to be

$$\frac{E_{out}}{E_{in}} = \frac{1+2RC_2s+R^2C_1C_2s^2}{1+R(C_1+2C_2)s+R^2C_1C_2s^2} \qquad (5)$$

Comparing the transfer functions of both networks, equations 1 and 5, it is obvious that they have similar characteristics. If R, C_1, C_2 (type *Ia*), equation 4, are replaced by C, R_2 and R_1 respectively, equation 4 becomes the transfer function for type *Ib* as shown by equation 1. A single graph giving the pole-zero laws for both types of network in normalized coordinates would be a convenient design tool. Such a graph is developed in the next section to allow the rapid location

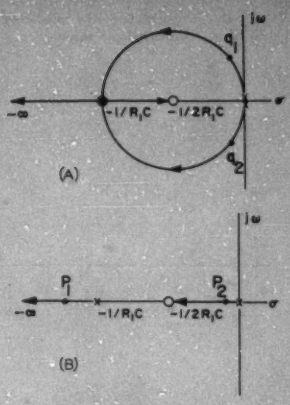


Fig. 2. Root loci of zeros and of poles of transfer function versus loop gain $1/R_2$

A—Of zeros
B—Of poles

of a pair of complex conjugate zeros anywhere in the left half-plane.

Normalized Characteristics

For design convenience, a graphical method is developed in this section. It should be made clear at the beginning that all derivations are given with no loading. It is desired to design the bridged-T networks with specified conjugate complex zeros. The poles are dependent on the zero locations. However, loading influences the pole-zero location of the transfer function. This effect will be discussed in the next section, where it will be used as part of the synthesis procedure.

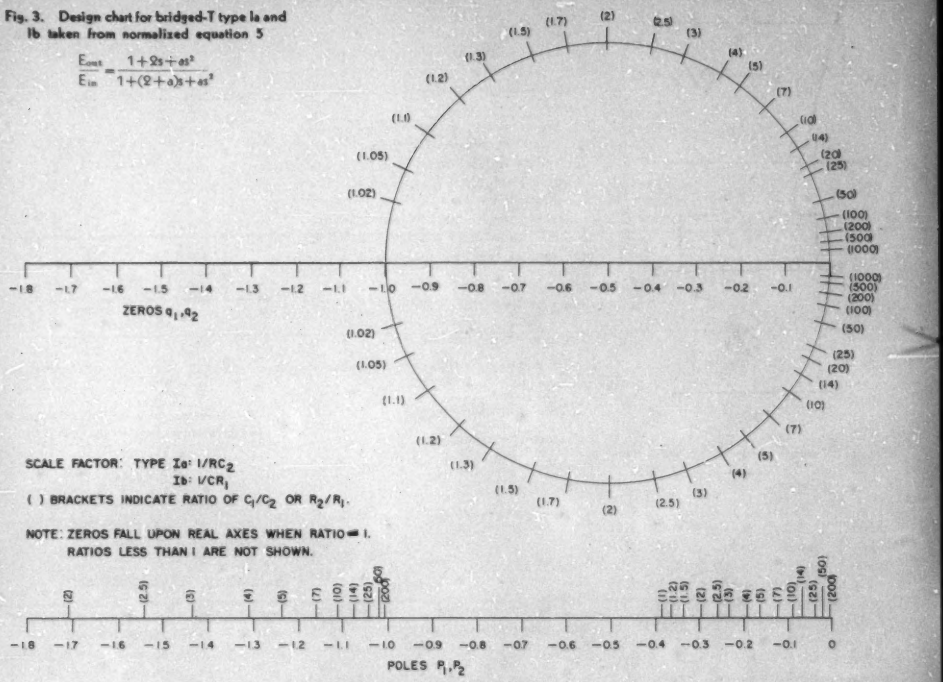
Normal Design Chart Development

As shown, type *Ia* and *Ib* networks are equivalent to each other. The same transfer function can be obtained just by switching their resistances and capacitances. Therefore, for the development of the design charts, only network *Ib* will be analyzed.

As shown in Fig. 2(A), the locus of complex zeros is along the circle centered at $-(1/2R_1C)$. If equation 2 is rewritten in terms of its damping ratio and undamped natural frequency (ω_n) as follows:

$$1+2R_1Cs+R_1R_2C^2s^2=\frac{1}{R_1R_2C^2}+\frac{2}{R_2C}s+s^2$$
$$=\omega_n^2+2\zeta\omega_n s+s^2=0 \qquad (6A)$$

then



undamped frequency $= \omega_n = \dfrac{1}{C\sqrt{R_1R_2}}$

damping ratio $= \zeta = \sqrt{R_1/R_2}$ (6B)

and

$\omega_n/\zeta = 1/R_1C$ (6C)

The damping ratio ζ is seen to depend upon the ratio, R_2/R_1. To normalize the locus of zeros, Fig. 2(A), a scale change can be introduced to cause the circle diameter to be equal to unity or, in other words, to cause $1/R_1C = 1$. With $R_1C = 1$, equation 2 is normalized and can be rewritten

$$\frac{1}{R_1R_2C^2} + \frac{2}{R_2C}s + s^2 = \frac{R_1}{R_2} + 2\frac{R_1}{R_2}s + s^2 = \omega_{nz}^2 + 2\zeta_{nz}\omega_{nz}s + s^2 = 0 \quad (7)$$

from which it may be seen that

normalized natural frequency

$$= \omega_{nz} = \sqrt{\frac{R_1}{R_2}} = \frac{1}{\sqrt{a}} \quad (8A)$$

normalized damping ratio

$$= \zeta_{nz} = \sqrt{R_1/R_2} = \frac{1}{\sqrt{a}} \quad (8B)$$

if it is assumed that $a = R_2/R_1$.

But from equations 6(B), 8(A) and 8(B), it is noticed that the original damping ratio, the normalized damping ratio, and the normalized natural frequency of oscillation are all equal.

$$\zeta = \zeta_{nz} = \omega_{nz} = \sqrt{R_1/R_2} = \frac{1}{\sqrt{a}} \quad (9)$$

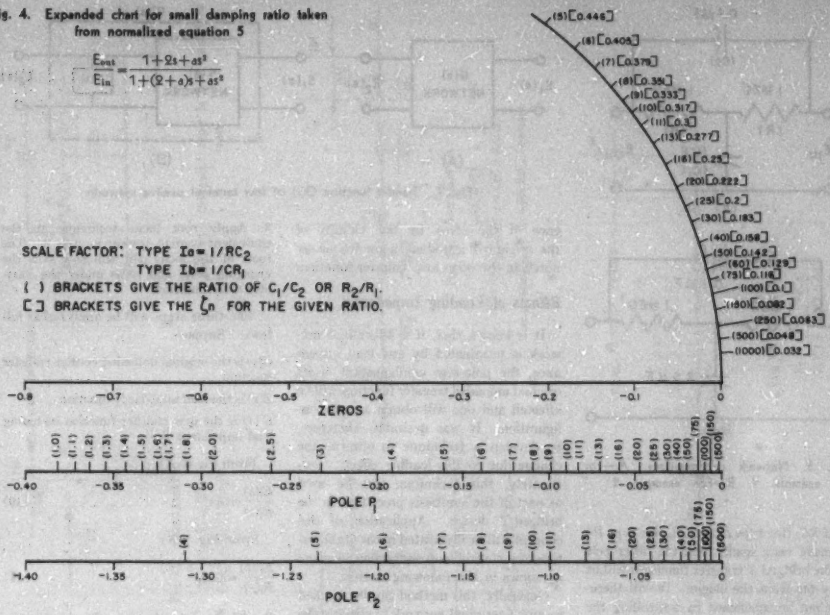
For complex conjugate zeros, it is necessary for R_2 to be greater than R_1. Normalizing equation 2 into equation 7 by setting $1/R_1C = 1$, the locus of zeros can be plotted, Fig. 3, as the ratio of R_2/R_1 is varied from unity to infinity. Similarly, equation 4 can also be normalized with the use of the scale factor R_1C and the locus of poles plotted in Fig. 3 as the ratio R_2/R_1 is varied from unity to infinity. Note that the plots could have been presented in terms of R_1/R_2, but it has been considered desirable to use large numbers rather than small decimals.

The ability of the bridged-T network to synthesize any left half-plane complex zero can be realized from Fig. 3. Here it can be seen that complex zeros can be readily placed with any desired damping ratio between zero and unity. Further since the damping ratio is invariant with scale change, the designer has only to change scale by selecting the proper value of $1/R_1C$ to adjust the complex zeros to any desired values on the required ζ line.

The chart of Fig. 3 can be used for networks of both types *Ia* and *Ib*. The numbers in the bracket indicate the ratio C_1/C_2 for network *Ia* or R_2/R_1 for network *Ib*. Direct readings of poles and zeros can be obtained only under the condition that $1/RC_2$ (network *Ia*) or $1/CR_1$ (network *Ib*) equal unity. The exact location of the poles and zeros can be obtained by multiplying the chart reading by $1/RC_2$ or $1/CR_1$. These factors are therefore the mentioned scale factors for the determination of the actual pole-zero locations. On the other hand, the desired locations of the zeros might be given and the problem would then be one of determining the required network. In this case, the given value of ζ would produce R_2/R_1 or C_1/C_2 from the chart. The ratio between the actual co-ordinate of the zeros and the normalized co-ordinates

Fig. 4. Expanded chart for small damping ratio taken from normalized equation 5

$$\frac{E_{out}}{E_{in}} = \frac{1+2\zeta s+as^2}{1+(2+a)s+as^2}$$



would produce the scale factor which gives $1/RC_2$ or $1/R_1C$.

For compensation in many practical applications, one must place the complex zeros close to the imaginary axis. To obtain more accuracy, an expanded chart of the region in the vicinity of the imaginary axis has been provided in Fig. 4.

Design Chart Application

Two examples of the application of the bridged-T charts are given:

Example 1: To determine of the voltage transfer function of the network shown in Fig. 5(A). This is a type *Ia* network with $1/RC_2=10$, and $C_1/C_2=10$. $1/RC_2$ becomes the scale factor. From the chart in Fig. 4:

Zeros: $-0.1\pm j0.3$
Poles: -0.09, -1.11

Multiply singularities by $1/RC_2$ to obtain the correct transfer function.

$$\frac{E_{out}}{E_{in}} = \frac{(s+1+j3)(s+1-j3)}{(s+0.9)(s+11.1)}$$

Example 2: To design a bridged-T type *Ia* network to give a pair of complex zeros at $\zeta=0.2$ and $\omega_n=2$, and then find find its voltage transfer function.

From equation 6(B): If $\zeta=0.2$, then R_2/R_1 or $C_1/C_2=25$

From the expanded chart, Fig. 4, for $C_1/C_2=25$

Zeros: $-0.04\pm j0.195$
Poles: -0.0384, -1.042
From equation 6(C),
$1/RC_2=\omega_n/\zeta=2/0.2=10$
Actual pole-zero location:
Zeros: $-0.4\pm j1.95$
Poles: -0.384, -10.42

Transfer function

$$=\frac{(s+0.4+j1.95)(s+0.4-j1.95)}{(s+0.384)(s+10.42)}$$

Network type *Ia* can be obtained by selecting its components under the condition that $C_1/C_2=25$ and $1/RC_2=10$. If one selects $R=1$ megohm, then $C_2=0.1$ microfarad and $C_1=2.5$ microfarads. This network is shown in Fig. 5(B).

Summary of Properties of Unloaded Bridged-T

From the previous analysis, the properties of both types of networks without loading may be summarized as follows. Refer to Fig. 1(A) and 1(B):

1. Product of poles = product of zeros (see equations 1 and 5)
2. Poles are always on negative real axis.
3. Zeros can be made negative real or complex according to the ratio of C_1/C_2 (*Ia*) or R_2/R_1 (*Ib*).

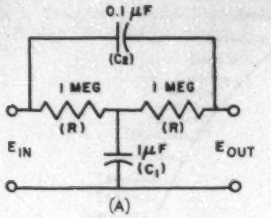
   Network *Ia:*
   $C_1/C_2>1$, complex
   $C_1/C_2=1$, double, negative real
   $C_1/C_2<1$, different, negative real

   Network *Ib:*
   $R_2/R_1>1$, complex
   $R_2/R_1=1$, double, negative real
   $R_2/R_1<1$, different, negative real

4. One pole is always closer to the origin than the real part of the zeros.
5. On the complex plane, the locus of zeros is a circle.
6. No d-c attenuation.
7. For very large C_1/C_2 or R_2/R_1, one pole and two zeros approach the origin; another pole approaches $-1/RC_2$ for type *Ia* and $-1/CR_1$ for type *Ib*.

Application

The property *g* is sometimes very useful in stabilization of a certain type of d-c servos, such as those positioning types with an inner loop pole at the origin and two more poles on the negative real axis. Suppose the open-loop transfer function of this system is $K/[s(1+\tau_1 s)(1+\tau_2 s)]$ In this case, the pole at the origin will associate with the two complex zeros and the one negative real pole which are all close to the origin; see Fig. 6.



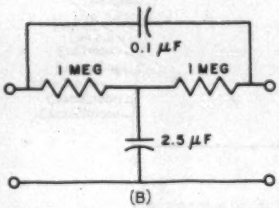


Fig. 5. Network configurations. A—For example 1. B—For example 2

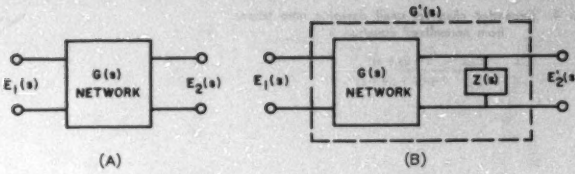


Fig. 7. Transfer function G(s) of four terminal passive networks

If RC_2 (for type *Ia*) or CR_1 (for type *Ib*) is made very small, then the other pole of the bridged-T transfer functions will be very far from the origin. It will therefore not be significant in controlling the system roots. Those singularities which group together very close to the origin will have little effect on the two remaining poles $-1/\tau_1$ and $-1/\tau_2$. Consequently the root locus with a break between these two remaining poles will be practically vertical and one can increase the gain to a very high level while maintaining a large damping exponent for the predominant roots. The main limitation for this application of the bridged-T is the roots near the origin. However, if the gain can be made very high with the system stability still satisfied, then the residues due to the roots close to the origin will be very small, and will contribute very little to the closed-loop transient response. This example, however, is somewhat oversimplified and must be used with caution. The presence of the zeros in the vicinity of the origin will introduce a low-frequency notch in the open loop transfer function.

Effects of Loading Impedance

It is known that, if a 4-terminal network is terminated by any load impedance, the pole-zero configuration of its original unloaded transfer function will be affected and one will obtain a new configuration. It was desirable, therefore, to develop a technique to observe the change due to this loading effect. Fortunately, this technique can be used as part of the synthesis procedure in the bridged-T design. Application of this concept will be illustrated in the stabilization of a control loop containing airplane dynamics in the following section.

Generally, this method can be applied to any 4-terminal network to observe the locus of the pole-zero change of the original transfer function when a variable load impedance is connected across its output terminals. It can be used, as well, for adjustment of the configuration.

Theory

The method employs Thevenin's theorem in conjunction with root locus. The pole-zero change of the network transfer function due to loading can be obtained by the following steps:

1. Having determined the original unloaded system transfer function, use Thevenin's theorem to find a new transfer function with load impedance $Z(s)$ being placed across the output terminal.
2. Convert this function into an equivalent feedback control system.
3. Apply root locus technique to the equivalent control system in step 2. The roots of the locus plot will now slow the changing poles and zeros under the variable load conditions.

The three steps will be analyzed as follows: Suppose

$G(s)$ is the original unloaded voltage transfer function.

$Z(s)$ is the load impedance function.

$G'(s)$ is the new transfer function including load impedance $Z(s)$.

From Fig. 7(A)

$$\frac{E_2(s)}{E_1(s)} = G(s) \tag{10}$$

From Fig. 7(B)

$$\frac{E_2'(s)}{E_1(s)} = G'(s) \tag{11}$$

By Thevenin's theorem, Fig. 7(B) can be transformed to Fig. 8. The new transfer function is

$$G'(s) = \frac{E_2'(s)}{E_1(s)} = \frac{G(s)Z(s)}{Z'(s)+Z(s)} \tag{12}$$

where $Z'(s)$ = impedance looking back into the output of network $G(s)$ with the input short-circuited.

Suppose $G(s)$, $Z(s)$, $Z'(s)$ are rewritten as the quotient of polynomials as follows:

$$G(s) = \frac{G_a(s)}{G_b(s)}$$

$$Z(s) = \frac{Z_a(s)}{Z_b(s)}$$

$$Z'(s) = \frac{Z_a'(s)}{Z_b'(s)} \tag{13}$$

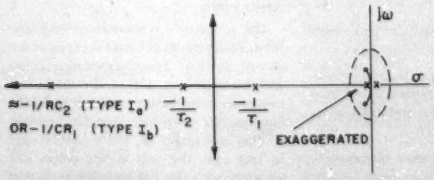


Fig. 6. Root locus plot of type I d-c servo with bridged-T network at extreme conditions as its compensating device

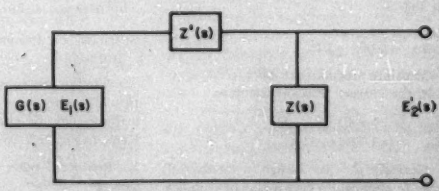


Fig. 8. Transformed circuit by Thevenin's theorem

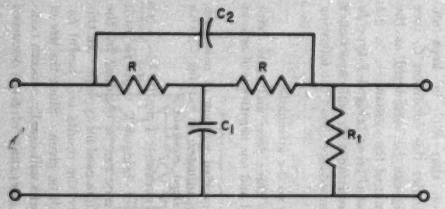


Fig. 9. Bridged-T type Ia with load resistance R_1

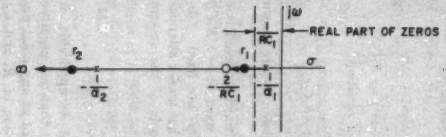


Fig. 10. Root loci for

$$\frac{2R}{R_1}\,\frac{(1+RC_1s/2)}{(1+\alpha_1 s)(1+\alpha_2 s)}$$

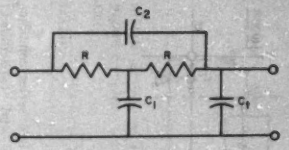


Fig. 11. Bridged-T type Ia with capacitive load

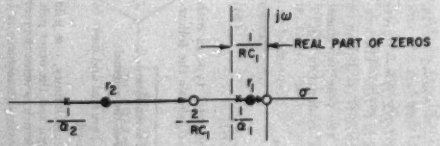


Fig. 12. Root loci of

$$2RC_1\,\frac{s(1+RC_1s/2)}{(1+\alpha_1 s)(1+\alpha_2 s)}$$

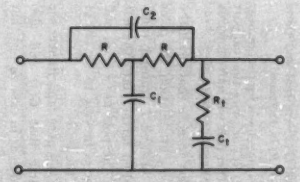


Fig. 13. Bridged-T type Ia with $Z(s) = R_1 + 1/C_1 s$

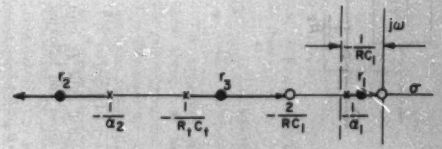


Fig. 14. Root loci for

$$2RC_1\,\frac{s(1+RC_1s/2)}{(1+\alpha_1 s)(1+\alpha_2 s)(1+R_1C_1 s)}$$

as a function of C_1 with R_1C_1 constant

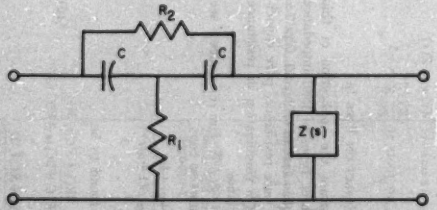


Fig. 15. Bridged-T type Ib with load impedance Z(s)

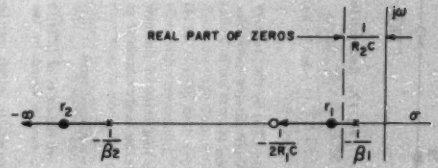


Fig. 16. Root loci for

$$\frac{R_2}{R_1}\,\frac{(1+2R_1C_2s)}{(1+\beta_1 s)(1+\beta_2 s)}$$

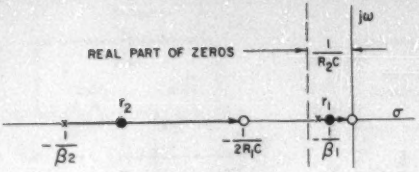


Fig. 17. Root loci for

$$C_tR_t\frac{s(1+2R_tC_ts)}{(1+\beta_{1s}s)(1+\beta_{2s}s)}$$

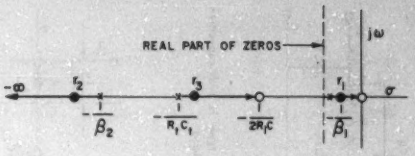


Fig. 18. Root loci for

$$C_tR_t\frac{s(1+2R_tC_ts)}{(1+\beta_{1s}s)(1+\beta_{2s}s)(1+R_tC_ts)}$$

Note that $Z_a'(s)=G_b'(s)$ by the property of all four terminal passive networks. Then, from equation 12

$$G'(s)=\frac{G(s)}{1+Z'(s)/Z(s)} \tag{14}$$

Then the poles and zeros of the new function $G'(s)$ can be easily obtained as follows:

1. Zeros of $G'(s)$ = combined zeros of $G(s)$ and $Z(s)$.
2. Poles of $G'(s)$ are the roots of $[Z'(s)]/[Z(s)]=-1$, which can be found by the root locus technique.
3. D-c attenuation $=G'(0)$ which can be obtained from equation 14.

Finally, $G'(s)$ can be written as

$$G'(s)=G'(0)\frac{(1-s/a_1)(1-s/a_2)\cdots(1-s/a_n)}{(1-s/r_1)(1-s/r_2)\cdots(1-s/r_n)} \tag{15}$$

where

$a_1, a_2, \ldots a_n$ are zeros of $G(s)$ and $Z(s)$

$r_1, r_2, \ldots r_n$ are roots of $\dfrac{[Z'(s)]}{[Z(s)]}=-1$

Loading Effect on Type Ia

To illustrate the application of this method, three types of load impedance will be applied first to bridged-T type *Ia*. These are pure resistance, pure capacitance, and a combination of resistance and capacitance.

Case Ia: $Z(s)$ is pure resistance; see Fig. 9. Suppose

$$G(s)=\frac{1+as+bs^2}{(1+\alpha_1s)(1+\alpha_2s)}$$

$$Z(s)=R_t$$

$Z'(s)$ is found to be

$$Z'(s)=\frac{2R(1+RC_1s/2)}{1+R(C_t+2C_1)s+R^2C_tC_1s^2} \tag{16}$$

or

$$Z'(s)=\frac{2R(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}$$

$$\frac{Z'(s)}{Z(s)}=\frac{2R}{R_t}\frac{(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}=-1 \tag{17}$$

A sketch of the root loci for poles of $G'(s)$ is shown in Fig. 10 by equating $Z'(s)/Z(s)=-1$.

Finally,

$$G'(s)=\frac{R_t}{2R+R_t}\frac{1+as+bs^2}{(1-s/r_1)(1-s/r_2)} \tag{18}$$

Analysis of Bridged-T type *Ia* with load resistance R_t shows the following properties:

1. D-c attenuation $=R_t/(2R+R_t)$.
2. Zeros are the same as original.
3. Both poles of $G'(s)$ are real and these move to the left if R_t is decreased.
4. Limit of poles movement for the one near the origin is $-2/RC_1$ or twice the real part of zeros. The other pole will move towards infinity along the negative real axis.

It is of interest to note that, with the introduction of d-c attenuation, the poles of the original bridged-T transfer function can be pulled farther into the left half-plane with the pole at $-1/\alpha_1$, and are even capable of moving to the left of the real part of the zeros.

Case IIa: $Z(s)$ is pure capacitance; see Fig. 11. Assume that $G(s)$ and also $Z'(s)$ are the same as in case *Ia* but that $Z(s)=1/C_ts$. Then

$$\frac{Z'(s)}{Z(s)}=2RC_t\frac{s(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}=-1 \tag{19}$$

The root loci of $Z'(s)/Z(s)=-1$ gives the locations of poles in $G'(s)$. The locations of the new poles of $G'(s)$ depend on C_t if $G(s)$ is fixed. The pole movement versus the loop gain C_t is shown in Fig. 12.

$G'(0)=1$

Then

$$G'(s)=\frac{1+as+bs^2}{(1-s/r_1)(1-s/r_2)} \tag{20}$$

The analysis of bridged-T type *Ia* with pure capacitive load C_t shows the following properties:

1. It contains no d-c attenuation.
2. Zeros are the same as the original.
3. Both poles of $G'(s)$ are real and move to the right if C_t is increased.
4. When C_t approaches infinity, one pole will approach the origin and the other pole approaches $-2/RC_1$, which is twice the real part of zero.

Case IIIa: $Z(s)$ is a series combination of resistance and capacitance; see Fig. 13. $G(s)$ and $Z'(s)$ are taken the same as the previous two cases.

$$Z(s)=\frac{1+R_tC_ts}{C_ts}$$

Zeros of $G'(s)=(1+as+bs^2)(1+R_tC_tS)$

Poles of $G'(s)$ = roots of equation

$$2RC_t\frac{s(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)(1+R_tC_ts)}=-1 \tag{21}$$

Root loci for expression 19 are shown in Fig. 14.

If R_t and C_t are arbitrarily selected, one zero in $G'(s)$ will be fixed at $-1/R_tC_t$; also, the poles of $G'(s)$ are fixed. Obviously, this network has no d-c attenuation. Suppose the roots of the system are fixed at r_1, r_2, and r_3; then

$$G'(s)=\frac{(1+as+bs^2)(1+R_tC_ts)}{(1-s/r_1)(1-s/r_2)(1-s/r_3)} \tag{22}$$

To make this analysis applicable, it should be emphasized that one negative real zero may be added at any point depending on the selection of R_t and C_t. Poles can be adjusted by varying C_t with the condition that R_tC_t be kept unchanged. Under this condition C_t (see equation 21) can be considered as the open-loop gain of the system.

Analysis of $G'(s)$ of bridged-T type *Ia* with series resistance R_t and capacitance C_t as load-impedance shows the following properties:

1. There is no d-c attenuation
2. Contains three poles and three zeros (two original and one negative real zero at $-1/R_tC_t$).
3. For a specified position of new zeros poles can be adjusted by varying C_t with R_tC_t maintained constant.

Loading Effects on Type Ib

Analysis of bridged-T type *Ib* with load impedance can be obtained by the same technique used for type *Ia*. In this section, only the results and useful expressions will be given. Since the use of these networks will usually concern complex zeros, the zeros of $G(s)$ discussed

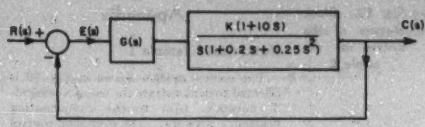


Fig. 19. Control system for example 1

in this section are assumed to be conjugate complex; see Fig. 15.
Let

$$G(s)=\frac{1+a's+b's^2}{(1+\beta_1 s)(1+\beta_2 s)} \tag{23}$$

All β_1, β_2, a', and b' are fixed and will be used for all cases in later discussion in this section.

Note that the real part of the zeros of $G(s)=1/R_2C$.

$$Z'(s)=\frac{R_2(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)} \tag{24}$$

Case Ib: $Z(s)$ is pure resistance (R_t). Zeros of $G'(s)$ = zeros of $G(s)$, and poles of $G'(s)$ = roots of

$$\frac{R_2}{R_t}\,\frac{(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)}=-1 \tag{25}$$

See Fig. 16 for root loci of equation 25. If R_t is fixed and the roots are at r_1 and r_2, then

$$G'(s)=\frac{R_t}{R_2+R_t}\,\frac{(1+a's+b's^2)}{(1-s/r_1)(1-s/r_2)} \tag{26}$$

Case IIb: $Z(s)$ is pure capacitance C_t. Zeros of $G'(s)$ = zeros of $G(s)$, and poles of $G'(s)$ = roots of

$$C_tR_2\,\frac{s(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)}=-1 \tag{27}$$

See Fig. 17 for root loci of equation 27. If C_t is fixed, the roots are at r_1 and r_2, then

$$G'(s)=\frac{(1+a's+b's^2)}{(1-s/r_1)(1-s/r_2)} \tag{28}$$

Case IIIb: $Z(s)$ is the series combination of resistance (R_t) and capacitance (C_t). Zeros of $G'(s)=(1+as+bs^2)(1+R_tC_ts)$, and poles of $G'(s)$ = roots of

$$C_tR_2\,\frac{s(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)(1+R_tC_ts)}=-1 \tag{29}$$

See Fig. 18 for root loci of equation 29. If R_t and C_t fix the roots at r_1, r_2, and r_3, then

$$G'(s)=\frac{(1+a's+b's^2)(1+R_tC_ts)}{(1-s/r_1)(1-s/r_2)(1-s/r_3)} \tag{30}$$

The properties of loading effect in these three cases for network type *Ib* are generally similar to that of type *Ia*. The differences are only the d-c attenuation terms and the limits which the roots of $[Z'(s)]/(Zs)]=-1$ will approach when the loop gain is increased towards infinity. Even in case *IIIb*, after another zero has been added, the poles must be adjusted (as in case *IIIa*) by varying C_t with R_tC_t maintained constant.

Use of Loading Effects

The behavior of the voltage transfer functions of both types of bridged-T networks under the three loading conditions have been discussed in the last two sections. One should note that, under the same loading condition, the two types of network behave in a similar manner. The only slight difference is the position of one zero in the equivalent open-loop transfer function $Z'(s)/Z(s)$ for the determination of poles in $G'(s)$. This zero will give a different termination of one root.

For example, in case *Ia*, this zero is at $-2/RC_1$ (see equation 17 and Fig. 10) and becomes the terminating zero for the root locus from the pole $-1/\alpha_1$. This is twice the real part of complex zeros of $G(s)$. In case *Ib*, (see equation 23 and Fig. 15) this zero is at the center of the circle passing through complex zeros of $G(s)$ and the origin. Consequently, even though $G(s)$ for networks *Ia* and *Ib* are the same, one will not obtain the same pole locations in $G'(s)$ for equivalent load resistance R_t.

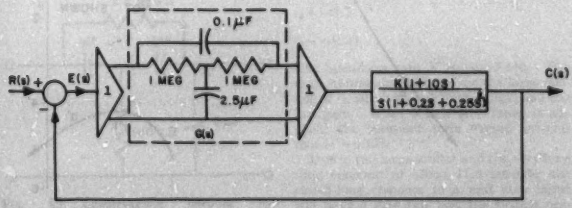


Fig. 21. Control system with bridged-T as compensating network G(s)

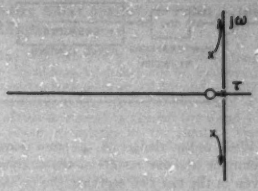


Fig. 20. Sketch of root loci for

$$\frac{K(1+10s)}{s(1+0.2s+0.25s^2)}$$

It has been shown that the damping ratio ζ of the zeros in $G(s)$ is controlled by R_2/R_1 for type *Ib* and C_1/C_2 for type *Ia*. If the $G(s)$ in both types are the same, then with very small ζ the limit to which the pole $-1/\beta_1$ (type *Ib*) can move away to catch a zero at $-1/(2R_1C)$ is greater than that to which the pole $-1/\alpha_1$ (type *Ia*) can move to catch a zero at $-2/RC_1$. Thus, if ζ is large, then type *Ia* would be preferred.

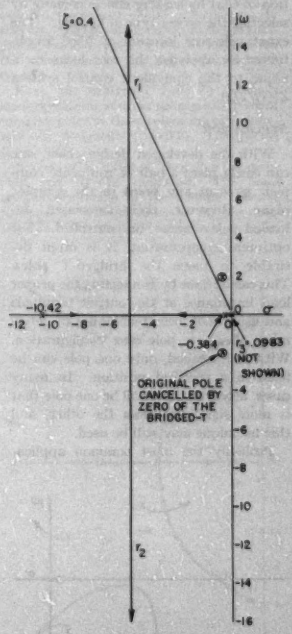


Fig. 22. Root loci of

$$\frac{40K(s+0.1)}{s(s+0.304)(s+10.42)}$$

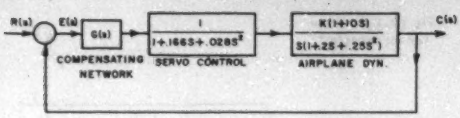


Fig. 23. Block diagram of the system for example 2

With a knowledge of the root locus technique and the results of the analysis given in the last two sections, the proper type of bridged-T can be selected to meet system requirements. Generally, the effect of loading can be used to adjust the pole positions of the transfer function of the bridged-T. However, while it is impossible to fix both poles at will, one can fix the one pole which is most significant. The capacitive load for bridged-T type *Ib* is used for pole adjustment in the numerical example 2 in the Appendix.

In actual applications, the bridged-T without load should be designed first by consultation of the design charts. Zeros of this network are fixed at the desired positions; then the complete pole-zero configuration of $G(s)$ is obtained. Modification of $G(s)$ by loading effects is made by selecting the proper type of loading. The exact pole-zero location of $G'(s)$ is obtained by applying the root locus technique to the equivalent control system $Z'(s)/Z(s)$.

Application

With the developed design chart, one can easily place a pair of conjugate complex zeros at any point in the complex plane. However, their associated unloaded poles cannot be controlled. For optimum compensation it is often desirable to move the bridged-T poles. This can be done by connecting the proper load impedance at the output terminals and using root locus techniques to determine the new pole-zero configuration. With this method, only one pole can be fixed in a specified position. In many cases, however, there will be one pole that is more significant than the other, and this technique may still be used.

Probably the most common application of bridged-T networks to d-c servo systems would be in those systems in which the transfer functions contain complex poles. The simplest way to compensate these types would be use of the cancellation technique of putting zeros on top of the complex poles. The system to be employed for illustration is the airplane auto pilot. Two examples are given in the Appendix. In example 1, only the bridged-T without loading effect will be used, but in example 2, the loading effect will be applied as well.

It should be stated that these examples are mainly intended to illustrate techniques for improving the system stability. If all specifications of the system are not entirely met, additional measures would, of course, be used. It should also be recognized that in actual systems, the exact pole-zero locations may not be exactly known and will often vary with change in environmental conditions. Exact cancellation of terms is, therefore, seldom if ever achieved.

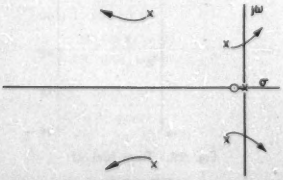


Fig. 24. Sketch of root loci of equation 35

Appendix

Numerical Example 1

The control system shown in Fig. 19 is selected to demonstrate the use of a bridged-T network. $G(s)$ is the compensating bridged-T network. The open-loop control transfer function, in this case, represents the short-period longitudinal mode of an airplane. The fin control servo is assumed perfect to avoid its effects on system performance. Without compensation, or $G(s)=1$, the root loci of this system are shown in Fig. 20. Without compensation, because of the small damping ratio, the system transient overshoot will be high.

With Bridged-T as Compensating Network

Pole cancellation will be used to improve the stability of the system. The complex poles in the original transfer function have a damping ratio of 0.2 and a natural frequency of 2 radians per second. In applying the cancellation method, the bridged-T network is designed to give complex zeros at $\zeta=0.2$ and $\omega_n=2$. Generally, either type of bridged-T can be used. For this particular case, the bridged-T type *Ia* has already been synthesized in the previous example 2 and will be used as the compensating network $G(s)$. The complete diagram of this control system is shown in Fig. 21. From the previous result

$$G(s)=\frac{(s+0.4+j1.95)(s+0.4-j1.95)}{(s+0.384)(s+10.42)}$$

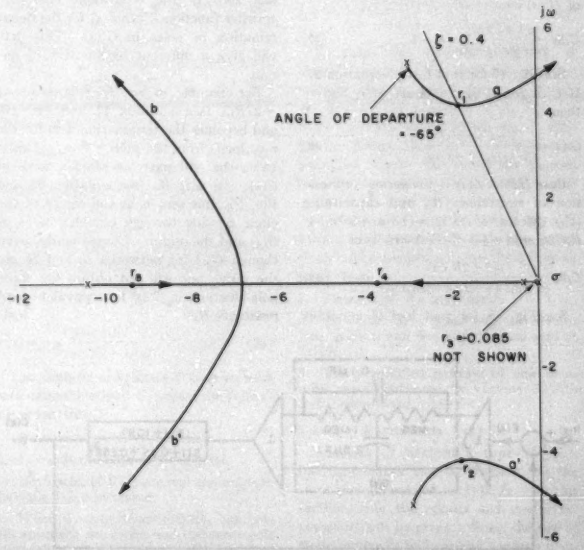


Fig. 25. Root loci of the system compensated by bridged-T without load

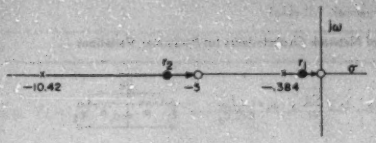


Fig. 26. Sketch of root loci for poles movement versus C_1

or

$$G(s)=\frac{(1+0.2s+0.25s^2)}{(1+2.7s)(1+0.096s)} \quad (31)$$

$$F_0(s)=\frac{K(1+10s)}{s(1+2.7s)(s+10.42)} \quad (32)$$

where $F_0(s)$ is the new open-loop transfer function, or

$$F_0(s)=\frac{40K(s+0.1)}{s(s+0.384)(s+10.42)} \quad (33)$$

Root loci of this function are shown in Fig. 22.

In this particular case, with a bridged-T as compensating network, the new root loci show more satisfactory results. The loci are nearly vertical, as the zero at −0.1 and the pole at the origin are very close together when compared with the other singularities.

With $\zeta=0.4$

$K=4.28$

$r_1, r_2=-5.24\pm j12.1$

$r_3=-0.0983$

With $r(t)$ = unit step input, residue at

$r_1=0.537\,|\underline{156°}$

$r_2=0.537\,|\underline{-156°}$

$r_3=-0.0172$

Then

$$c(t)=1-0.0172e^{-0.0983t}+1.074e^{-5.24t}\cos(12t+156°) \quad (34)$$

With $\zeta=0.4$, the overshoot from r_1 and r_2 $\cong 25\%$[3]. In this case, the residue of the root r_3 (with unit step input) is negative and very small, so the overshoot will be slightly less than 25%. Actually, due to the very long time constant of r_3, it will give a long-tail effect to the closed loop performance. However, the residue at this root is so small that, in practice, one can neglect it.

It should be pointed out that this analysis is made on the basis of exact cancellation which, of course, is impossible. The technique can still be applied for if the zeros are still near the poles, the residues at those roots will be very small and they will not effect the closed-loop performance significantly.

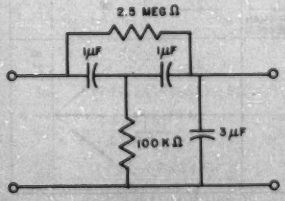


Fig. 27. Compensating network for example 2

Numerical Example 2

The application of bridged-T with loading effect will be illustrated in this example to improve the system performance. The same problem used in example 1 is selected for illustration but, in addition, another servo control unit is placed in cascade with the original. This system is shown in Fig. 23.

With $G(s)=1$

$$F_0(s)=\frac{K(1+10s)}{s(1+0.25+0.25s^2)\times(1+0.166s+0.028s^2)} \quad (35)$$

The root loci of this system are shown in Fig. 24. Without compensation, it is seen that the predominating roots of the system are about to move to the right half-plane, even at the very low gain. It is desired to design a compensating network $G(s)$ to improve the system stability. For this particular problem, the bridged-T networks again seem to to be very applicable.

Bridged-T Applied Without Load

If the bridged-T network, which has been used in the last problem, is applied to this system as a compensating network $G(s)$ to cancel the complex poles of the airplane function, the system stability is improved. See Fig. 25 for the system root loci plot.

If the damping ratio of 0.4 is selected for gain adjustment, the roots of the closed-loop system are

$r_1, r_2=-1.75\pm j4.15$

$r_3=-0.085$

$r_4=-3.7$

$r_5=-9.39$

The open-loop gain K equals 0.446. The most undesirable features are the large residue of r_3, which is close to the origin and the low gain. When a unit step input is applied, the transient term caused by this root is $-0.162e^{-0.085t}$.

Due to this large residue and the very long time constant of about 11.8 seconds, the closed-loop response to a unit step input will have a long tail, and the system performance is not considered satisfactory.

Loading Effect Applied for Poles Adjustment

The root loci plot in Fig. 25 shows that the closed-loop response can be improved by moving both poles at −0.384 and −10.42 towards the origin until the loci a and a' are about to change to the left (or b and b' are about to change to the right). The required location of the new poles can be roughly estimated by checking the angle of departure from the complex poles. If all singularities are held constant, the pole at −10.42 has to be moved to −6 to satisfy this condition. If a capacitance is used for loading the original bridged-T, both poles at −10.42 and −0.384 can be moved to the right.

Another improvement that may be considered at the same time is to try to cancel the zero at −0.1 by moving the pole from −0.384. If an approximate cancellation at this point can be made, the calculation is simplified and the effect of the long time constant transient term will be eliminated.

For this particular problem, the loading effect of case *IIb* seems very suitable. The bridged-T type *Ib* used with load capacitance will have terminating loci from both poles at −5 and the origin. If the case *IIa* is used, the loci will terminate at −0.8 and origin. In case *IIb*, there is a possibility of obtaining both zero cancellation and another pole moving to the vicinity of −6, while, in case *IIa*, the pole at −0.384 will change only slightly. The previous results from the analysis of loading case *IIb* will be applied for the adjustment of the poles of bridged-T type *Ib*.

First, the bridged-T type *Ib* (without load) containing zeros at the same location of complex poles in the airplane transfer function is designed: $\zeta_{n2}=0.2$; $1/R_1C=10$; $R_2/R_1=25$.

If $C=1\mu F$ is selected, then $R_1=0.1$ megohm and $R_2=2.5$ megohms. The pole-zero configuration of the voltage transfer function is still the same as for the network *Ia* which was used in the last example.

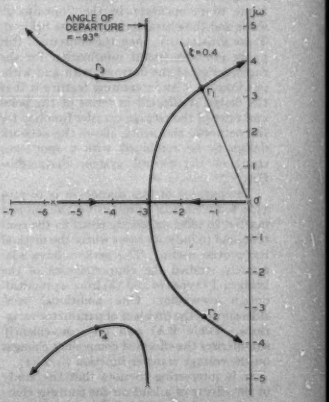


Fig. 28. Root loci of the system compensated by bridged-T type Ib with capacitive load

Next, using the capacitive load C_t and applying equation 27 to obtain the relation for pole adjustment yields

$$\frac{2C_s(s+5)}{(s+10.42)(s+0.384)} = -1 \qquad (36)$$

See Fig. 26 for the root loci of pole movement.

If the root r_1 is fixed at -0.1, then $C_t = 2.95$, and $r_2 = -5.75$. The network which is to be used for $G(s)$ is shown in Fig. 27. The root loci plot of the system is given in Fig. 28. With the damping ratio $= 0.4$, the results are $r_1, r_2 = -1.4 \pm j3.25$; $r_3, r_4 = -4.5 \pm j3.55$; $K = 2.04$.

If $r(t) =$ unit step input

$$c(t) = 1 + 1.57\epsilon_t^{-1.4t} \cos(3.25t + 96°) + 0.89\epsilon^{-4.5t} \cos(3.55t + 200°) \qquad (37)$$

The overshoot in this case is roughly estimated at about 26%.

References

1. Control-System Dynamics (book). **W. R. Evans.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 156–57.

2. Automatic Feedback Control System Synthesis (book). **John G. Truxal.** McGraw-Hill Book Company, Inc., 1955, pp. 38–43.

Discussion

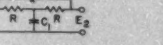
Theodore A. Savo (Hughes Aircraft Company, Culver City, Calif.): The authors are to be complimented on their interesting paper, which helps to bridge the gap between what might be called network synthesis and network design. The root-locus method, used in the graphical study of the poles and zeros of the bridged-T network in this paper, can of course be used for other networks the extension being straightforward. If the decision has been made to compensate in the "open-loop" sense and the characteristics of the bridged-T are satisfactory, then the content of this paper provides useful information for the rapid design of the network with and without loading. An interesting feature is that the design is affected in terms of the poles and zeros of the voltage transfer function for the network, and hence allows the network design to be combined with a root-locus study of the control system characteristics.

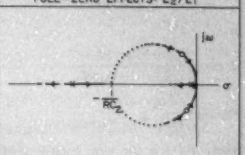
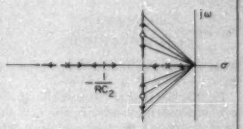
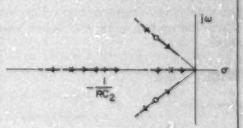
The purpose of this discussion is to provide some parallel opinions on the subject matter, to offer some extensions to the content, and to indicate cases where the method has proved useful. The authors have adequately studied the characteristics of the bridged-T (types *Ia* and *Ib*) from a practical-design viewpoint. One additional consideration is the problem of parameter variations. Table I(A) and (B) conveniently summarizes the effects of component changes on the voltage transfer function $E_2/E_1(s)$.

It is interesting to note that the study of the effects of a load on the transfer characteristics of the bridged-T can be conveniently expressed in terms of the short-circuit admittance parameters (y_{12}, y_{22}).

Table I. Summary of Network Characteristics for Parameter Variations

A. NETWORK TYPE Ia: $\zeta = \sqrt{\frac{C_2}{C_1}}$, $\omega_n = \frac{1}{R\sqrt{C_1 C_2}}$



PARAMETER	VARIATION	CHARACTERISTIC CHANGE	POLE–ZERO EFFECTS: E_2/E_1
C_1	INCREASED	$\frac{C_1}{C_2}$ – INCREASES ζ – DECREASES $\frac{1}{RC_2}$ – UNCHANGED	
	DECREASED	$\frac{C_1}{C_2}$ – DECREASES ζ – INCREASES $\frac{1}{RC_2}$ – UNCHANGED	
C_2	INCREASED	$\frac{C_1}{C_2}$ – DECREASES ζ – INCREASES $\frac{1}{RC_2}$ – DECREASES (FASTER THAN ζ INCREASES)	
	DECREASED	$\frac{C_1}{C_2}$ – INCREASES ζ – DECREASES $\frac{1}{RC_2}$ – INCREASES (FASTER THAN ζ DECREASES)	
R	BOTH INCREASED SIMULTANEOUSLY	$\frac{C_1}{C_2}$ – UNCHANGED ζ – CONSTANT $\frac{1}{RC_2}$ – DECREASES	
	BOTH DECREASED SIMULTANEOUSLY	$\frac{C_1}{C_2}$ – UNCHANGED ζ – CONSTANT $\frac{1}{RC_2}$ – INCREASES	

B. NETWORK TYPE Ib: $\zeta = \sqrt{\frac{R_1}{R_2}}$, $\omega_n = \frac{1}{C\sqrt{R_1 R_2}}$

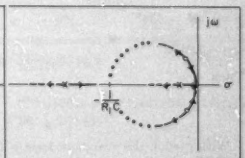
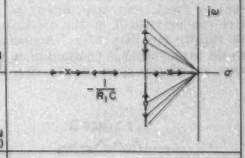
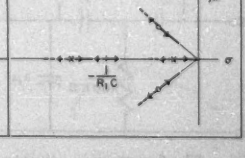
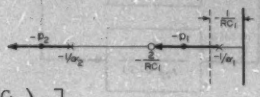
R_2	INCREASED	$\frac{R_2}{R_1}$ – INCREASES ζ – DECREASES $\frac{1}{R_1 C}$ – UNCHANGED	
	DECREASED	$\frac{R_2}{R_1}$ – DECREASES ζ – INCREASES $\frac{1}{R_1 C}$ – UNCHANGED	
R_1	INCREASED	$\frac{R_2}{R_1}$ – DECREASES ζ – INCREASES $\frac{1}{R_1 C}$ – DECREASES (FASTER THAN ζ INCREASES)	
	DECREASED	$\frac{R_2}{R_1}$ – INCREASES ζ – DECREASES $\frac{1}{R_1 C}$ – INCREASES (FASTER THAN ζ DECREASES)	
C	BOTH INCREASED SIMULTANEOUSLY	$\frac{R_2}{R_1}$ – UNCHANGED ζ CONSTANT $\frac{1}{R_1 C}$ – DECREASES	
	BOTH DECREASED SIMULTANEOUSLY	$\frac{R_2}{R_1}$ – UNCHANGED ζ – CONSTANT $\frac{1}{R_1 C}$ – INCREASES	

Table II. Summary of Loaded Types Ia and Ib Networks

(A) Network Ia

(a) $Y_L = 1/R_L$

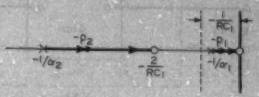


$$\frac{Y_L}{y_{22}}(S) = \frac{2R}{R_L}\left[\frac{\left(1+\frac{RC_1}{2}S\right)}{(1+\alpha_1 S)(1+\alpha_2 S)}\right]$$

$$G(S) = \frac{1+aS+bS^2}{(1+\alpha_1 S)(1+\alpha_2 S)}$$

$$G'(S) = \frac{R_L}{2R+R_L}\left[\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

(b) $Y_L = C_L S$

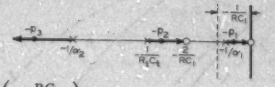


$$\frac{Y_L}{y_{22}}(S) = 2RC_L\left[\frac{S\left(1+\frac{RC_1}{2}S\right)}{(1+\alpha_1 S)(1+\alpha_2 S)}\right]$$

$$G(S) = \frac{1+aS+bS^2}{(1+\alpha_1 S)(1+\alpha_2 S)}$$

$$G'(S) = \frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}$$

(c) $Y_L = \left[\frac{C_L S}{R_L C_L S+1}\right]$



$$\frac{Y_L}{y_{22}}(S) = \frac{2RC_LS\left(1+\frac{RC_1}{2}S\right)}{(1+\alpha_1 S)(1+\alpha_2 S)(1+R_LC_LS)}$$

$$G(S) = \frac{1+aS+bS^2}{(1+\alpha_1 S)(1+\alpha_2 S)}$$

$$G'(S) = \frac{(1+aS+bS^2)(1+R_LC_LS)}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)\left(1+\frac{S}{p_3}\right)}$$

(d) $Y_L = \left[\frac{R_LC_LS+1}{R_L}\right]$

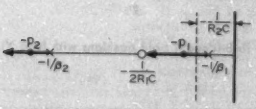
$$\frac{Y_L}{y_{22}}(S) = \frac{2R}{R_L}\left[\frac{\left(1+\frac{RC_1}{2}S\right)(1+R_LC_LS)}{(1+\alpha_1 S)(1+\alpha_2 S)}\right]$$

$$G(S) = \frac{1+aS+bS^2}{(1+\alpha_1 S)(1+\alpha_2 S)}$$

$$G'(S) = \frac{R_L}{R_L+2R}\left[\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

(B) Network Ib

(a) $Y_L = 1/R_L$

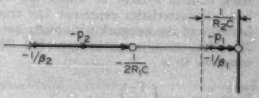


$$\frac{Y_L}{y_{22}}(S) = \frac{R_2}{R_L}\left[\frac{(1+2R_1CS)}{(1+\beta_1 S)(1+\beta_2 S)}\right]$$

$$G(S) = \frac{1+aS+bS^2}{(1+\beta_1 S)(1+\beta_2 S)}$$

$$G'(S) = \frac{R_L}{R_2+R_L}\left[\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

(b) $Y_L = C_L S$

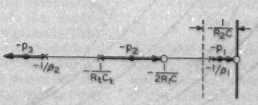


$$\frac{Y_L}{y_{22}}(S) = \frac{R_2C_LS(1+2R_1CS)}{(1+\beta_1 S)(1+\beta_2 S)}$$

$$G(S) = \frac{1+aS+bS^2}{(1+\beta_1 S)(1+\beta_2 S)}$$

$$G'(S) = \frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}$$

(c) $Y_L = \left[\frac{C_LS}{1+R_1C_LS}\right]$

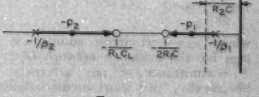


$$\frac{Y_L}{y_{22}}(S) = \frac{R_2C_LS(1+2R_1CS)}{(1+\beta_1 S)(1+\beta_2 S)(1+R_1C_LS)}$$

$$G(S) = \frac{1+aS+bS^2}{(1+\beta_1 S)(1+\beta_2 S)}$$

$$G'(S) = \frac{(1+aS+bS^2)(1+R_1C_LS)}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)\left(1+\frac{S}{p_3}\right)}$$

(d) $Y_L = \left[\frac{R_LC_LS+1}{R_L}\right]$



$$\frac{Y_L}{y_{22}}(S) = \frac{R_2}{R_L}\left[\frac{(R_LC_LS+1)(1+2R_1CS)}{(1+\beta_1 S)(1+\beta_2 S)}\right]$$

$$G(S) = \frac{1+aS+bS^2}{(1+\beta_1 S)(1+\beta_2 S)}$$

$$G'(S) = \frac{R_L}{R_2+R_L}\left[\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

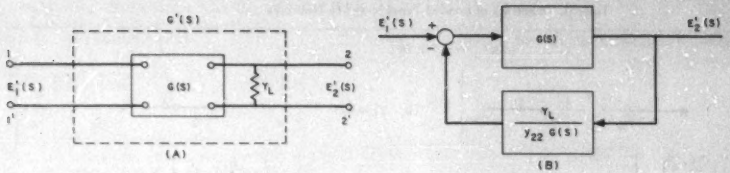


Fig. 29. Network loading and equivalent feedback system

The voltage transfer function for the network with no load is

$$\frac{E_2}{E_1}(S)=\frac{y_{12}}{y_{22}}\triangleq G(S) \tag{38}$$

(Since it would add little to the discussion, the minus sign normally associated with equation 38 has been arbitrarily omitted.)

The expression for the voltage transfer function if the network is loaded with an admittance Y_L can be expressed as

$$\frac{E_2}{E_1}(S)=\frac{y_{12}}{y_{22}+Y_L}\triangleq G'(S) \tag{39}$$

Slight algebraic manipulation of equation 39 yields

$$G'(S)=\frac{E_2}{E_1}(S)=\frac{y_{12}/y_{22}}{1+\left[\frac{Y_L}{y_{22}}\left(\frac{y_{22}}{y_{12}}\right)\right]\left(\frac{y_{12}}{y_{22}}\right)}\triangleq \frac{G(S)}{1+\left(\frac{Y_L}{y_{22}G(S)}\right)G(S)} \tag{40}$$

Equation 40 defines an equivalent feedback control system in which the closed loop characteristics represent the transfer function of the network with load; see Fig. 29. The open-loop gain from Fig. 29 and equation 40 is

$$\text{Loop gain}=-\left(\frac{Y_L}{y_{22}}\right) \tag{41}$$

Equation 41 can be correlated with the method used in the paper, namely

$$\frac{Y_L}{y_{22}}\triangleq\frac{Z'(S)}{Z(S)} \tag{42}$$

The characteristics of the loaded network $G'(s)$ can now be studied by the root-locus method, the characteristic equation being

$$1+\frac{Y_L}{y_{22}}(S)=0 \tag{43}$$

where y_{22} = short-circuit admittance function for the unloaded network $G(s)$; and Y_L = admittance of any arbitrary load. Specifically, the short-circuit admittance functions for the two types of bridged-T networks are, for type *Ia*:

$$y_{22}=\frac{1+R(C_1+2C_2)S+C_1C_2R^2S^2}{2R\left(1+\frac{RC_1}{2}S\right)} \tag{44}$$

and for type *Ib*:

$$y_{22}=\frac{1+C(R_2+2R_1)S+R_1R_2C^2S^2}{R_2(1+2R_1CS)} \tag{45}$$

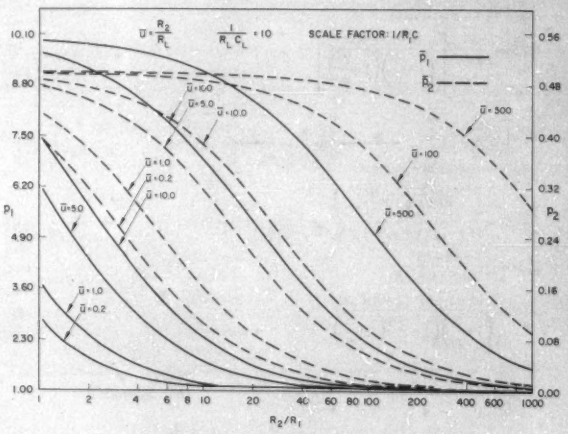


Fig. 30. Type Ib bridged-T pole positions with load $Y_L=(R_LC_L+1)/R_L$

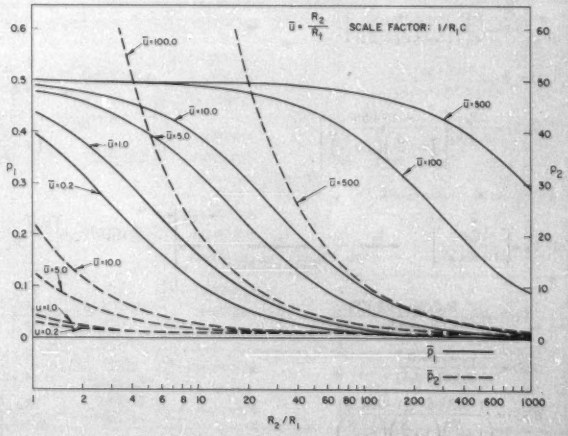


Fig. 31. Type Ib bridged-T pole positions with load $Y_L=1/R_L$

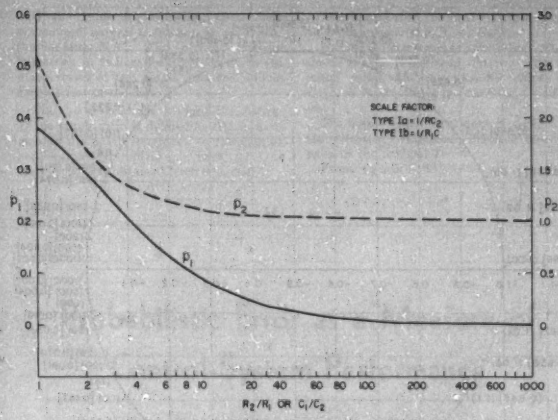


Fig. 32. Bridged-T pole positions, unloaded

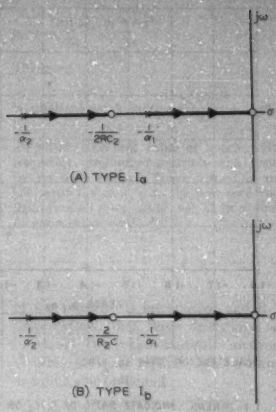


Fig. 34. Bridged-T pole locations with resistive source load R_s

The authors have studied the bridged-T characteristics with three types of simple loads. One additional type of simple load, not considered, is the case of a parallel combination of one resistor and capacitor. For this case the load admittance is

$$Y_L = \frac{R_L C_L S + 1}{R_L} \tag{46}$$

The function Y_L/y_{22} becomes, for type *Ia*:

$$\frac{Y_L}{y_{22}}(S) \triangleq \frac{Z'(S)}{Z(S)} = \frac{2R}{R_L}\,\frac{\left(1+\frac{RC_1}{2}S\right)(1+R_L C_L S)}{(1+\alpha_1 S)(1+\alpha_2 S)} \tag{47}$$

and for type *Ib*:

$$\frac{Y_L}{y_{22}}(S) \triangleq \frac{Z'(S)}{Z(S)} = \frac{R_1}{R_L}\,\frac{(1+2R_1 CS)(1+R_L C_L S)}{(1+\beta_1 S)(1+\beta_2 S)} \tag{48}$$

The zeros of the transfer function with the parallel R_L, C_L load, $G'(S)$, are, of course, the same as the zeros of the network transfer function without load $G(S)$. Both networks, however, have a d-c attenuation. A convenient summary of the loading effects on the pole locations is given in Table II(A) and (B) for the three cases considered in the paper and for the additional case just presented. The information shown in Table II is useful in determining qualitatively the effects of loading (intentional or unintentional); however, in a specific problem, if the pole locations are to be determined exactly, a design chart of the form shown in Fig. 30 is appropriate for engineers interested in rapid laboratory-type design. Fig. 30 depicts the type *Ib* network with the parallel, R_L, C_L type of load and the normalized time constant of the load set at $T_L = 0.1$ second. Fig. 31 illustrates the pole positions for the type *Ib* network with simple resistive loading R_L. Fig. 32 gives the pole positions for the unloaded bridged-T. It should be noted that the pole positions shown in Fig. 30 and 31 are computed with R_1C normalized to unity; hence a $1/R_1C$ scale factor is necessary. Similar comments apply to Fig. 32.

It is logical to consider the characteristics of the bridged-T network for the case of source loading as illustrated in Fig. 33.

The transfer function for the case of source loading is, for type *Ia*:

$$\hat{G}(S) = \frac{1+2RC_2 S+C_1C_2R^2S^2}{1+[R(C_1+2C_2)S+R_sC_1]S+RC_1C_2(R+2R_s)S^2} \tag{49}$$

and for type *Ib*:

$$\hat{G}(S) = \frac{1+2R_1CS+R_1R_sC^2S^2}{1+C(2R_1+R_s+2R_s)S+R_sC^2S^2} \tag{50}$$

As might be expected for resistive source loading, equations 49 and 50 reveal that the zeros of the network are not altered. Study of the pole migration using the root-locus method is best effected in the case of source loading by using the open-circuit impedance functions $z_{12}(S)$ and $z_{11}(S)$. The voltage transfer function of the network with no load is

$$\frac{E_2}{E_1}(S) = \frac{z_{12}(S)}{z_{11}(S)} \tag{51}$$

If an arbitrary source load Z_s is added, then the voltage transfer function becomes

$$\frac{E_2}{E_1}(S) = \frac{z_{12}}{z_{11}+Z_s} \tag{52}$$

Manipulation of equation 52, in a manner analogous to equation 40, yields

$$\frac{E_2}{E_1}(S) = \frac{z_{12}/z_{11}}{1+\frac{Z_s}{z_{11}}} \tag{53}$$

Equation 53 indicates that the poles of the transfer characteristic can be studied with a root-locus plot of Z_s/z_{11} in a manner analogous to Y_L/y_{22} for load impedances: see equation 43.

The open-circuit driving point impedance, $z_{11}(S)$, for the bridged-T network is, for type *Ia*:

$$z_{11} = \frac{1+R(C_1+2C_2)S+C_1C_2R^2S^2}{C_1S(1+2RC_2S)} \tag{54}$$

and for type *Ib*:

$$z_{11} = \frac{1+C(2R_1+R_2)S+R_1R_2C^2S^2}{2CS\left(1+\frac{R_2C}{2}S\right)} \tag{55}$$

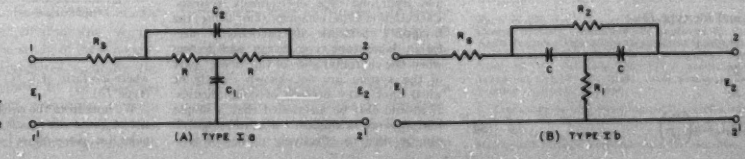


Fig. 33. Bridged-T network with source loading

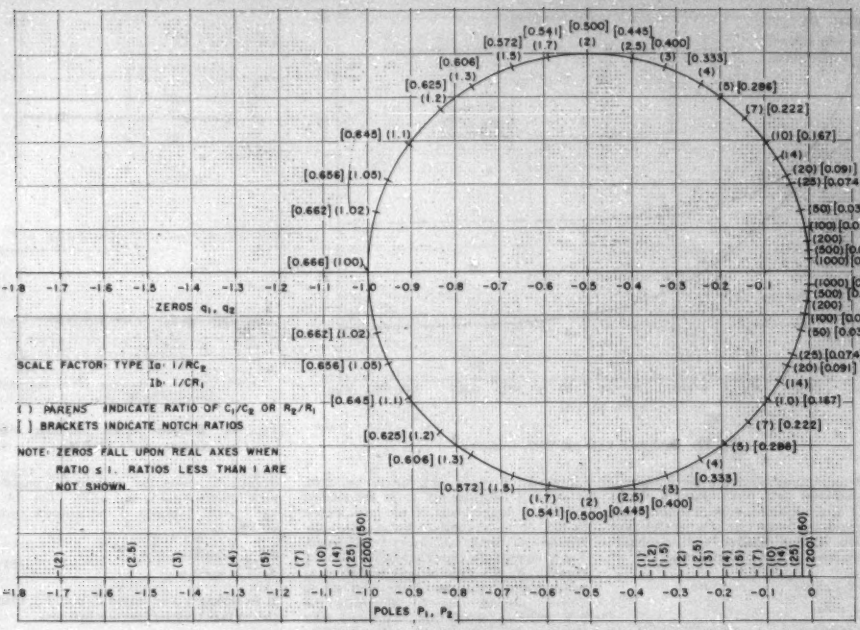


Fig. 35. Design chart for bridged-T type Ia and Ib taken from normalized equation 5

$$\frac{E_{out}}{E_{in}} = \frac{1+2s+as^2}{1+(2+a)s+as^2}$$

The resultant ratio $Z_s/z_{11}(S)$ is $(Z_1=R_s)$ for type *Ia:*

$$\frac{Z_s}{z_{11}}(S) = \frac{R_sC_1S(1+2RC_2S)}{(1+\alpha_1S)(1+\alpha_2S)} \tag{56}$$

and for type *Ib:*

$$\frac{Z_s}{z_{11}}(S) = \frac{2R_sCS\left(1+\frac{R_2C}{2}S\right)}{(1+\alpha_1S)(1+\alpha_2S)} \tag{57}$$

The parameters $1/\alpha_1$ and $1/\alpha_2$ in equations 56 and 57 characterize the pole locations for $R_s=0$.

Figure 34 depicts the pole migration derived from equations 56 and 57. With the frequency normalization used for the unloaded bridged-T, the gain calibration of the loci of Fig. 34 is quite simple, for example, for type *Ia:*

$$RC_2=1$$

$$C_1/C_2=n$$

$$R_sC_1=n\left(\frac{R_s}{R}\right) \tag{58}$$

and for type *Ib:*

$$R_1C=1$$

$$\frac{R_2}{R_1}=n$$

$$R_sC=n\left(\frac{R_s}{R_2}\right) \tag{59}$$

Consequently, for a given n value which fixes the complex zeros, the variation in the pole locations can be studied by varying the ratios (R_s/R) and R_s/R_2 respectively. In conclusion, it is obvious that if the network level can be designed to yield $R>>R_s$ or $R_2>>R_s$, then the unloaded design charts can be used. In cases where the source resistance is not negligible or drifts substantially, the results shown in Fig. 34 are useful.

The case of both input and output loading has also been considered by the aforementioned method; however, the number of variable parameters becomes large unless special symmetry is assumed.

In conclusion it is worth mentioning that the root-locus method of designing the bridged-T has been successfully used in effecting the compensation of both electric and hydraulic type of feedback systems. Cancellation-type compensation using the bridged-T network is particularly useful in laboratory-type design and testing when the characteristics of the "fixed" part of the system are not known exactly or when they vary because of nonlinearities. It should also be mentioned that a simple lead network in tandem with the bridged-T can be used to effectively move the low-frequency pole of the unloaded bridged-T to a higher frequency.

Propat Chandaket and **Allen B. Rosenstein:** The authors wish to thank Dr. Savo for his interesting discussion and his additions to the work of this paper. The use of short-circuit admittance parameters to study the loading effect is more convenient for the circuit designers. Dr. Savo's plots of pole positions for the case of bridged-T under no load and under loaded conditions of pure resistance and parallel *R-C* will be particularly useful for the designer.

Since the bridged-T is frequently employed as an a-c filter, the authors would like to extend the utility of the method developed a little further by adding the notch ratio and notch width characteristics of the unloaded bridged-T of both types.

These properties can be obtained quite easily from the normalized equation in Fig. 32.

$$\frac{E_{out}}{E_{in}} = \frac{1+2s+as^2}{1+(2+a)s+as^2} = \frac{\left(1+\frac{s}{q_1}\right)\left(1+\frac{s}{q_2}\right)}{\left(1+\frac{s}{p_1}\right)\left(1+\frac{s}{p_2}\right)} \tag{60}$$

where a = ratio of C_1/C_2 (type *Ia*) or R_2/R_1 (type *Ib*).

We note from the equation that the product of the zeros equals the product of the poles, i.e., $q_1q_2=p_1p_2=1/a$.

If the zeros are complex, then their normalized natural frequency is $\omega_{n2}1/\sqrt{a}$; see equation 8(A). This, of course, is the frequency in radians locating the center (in a log scale) of the notch, i.e., the double break point of the zeros.

The d-c gain of the bridged-T is unity. If the notch ratio is defined as the ratio of gain at the notch frequency to the gain at zero frequency, replacing s with $j(1/\sqrt{a})$ the following is obtained from equation 1:

$$\text{Notch ratio} = \frac{2}{2+a} \qquad (61)$$

This shows that the notch ratio depends only on the ratio of C_1/C_2 or R_2/R_1.

For design convenience, the design chart shown in Fig. 3 has been redrawn with the notch ratios for the various values of C_1/C_2 or R_2/R_1 added in brackets; see Fig. 35.

The notch width frequency is defined as the frequency difference between the two poles of the bridged-T.

From equation 60

$$p_1 = -\frac{(2+a)}{2a} + \frac{\sqrt{4+a^2}}{2a}$$

$$p_2 = -\frac{(2+a)}{2a} - \frac{\sqrt{4+a^2}}{2a}$$

This gives

$$\text{Normalized notch width} = \frac{\sqrt{4+a^2}}{a} \qquad (62)$$

To obtain the real value of notch width frequency, one must multiply the result by its scale factor. Since $p_1p_2 = 1/a$ and the zeros fall at $1/\sqrt{a}$, the pole-zero configuration on a log scale such as in a Bode diagram is symmetrical.

Probabilistic Error as a Measure of Control-System Performance

J. ZABORSZKY
MEMBER AIEE

J. W. DIESEL
ASSOCIATE MEMBER AIEE

AUTOMATIC FEEDBACK-control system synthesis occupies a rather unique place among the various fields of engineering due to the indirectness of its approach. An antenna will be designed for a certain radiation pattern because this radiation pattern is called for in the intended application. On the other hand, a servomechanism which is destined to handle step inputs exclusively may be synthesized by the techniques of Nyquist, Nichols, or Bode. This means that the system is actually designed for a frequency response: to exhibit certain resonance characteristics when subjected to steady-state sinusoidal inputs of varying frequency. Yet the system will never be subjected to this latter type of input. The justification of this approach is that the experience of the designer provides the basis for expecting that the system will perform properly in its actual input environment if it is built for a specific frequency response. Such examples of indirectness could be enumerated almost endlessly.

In brief, a common characteristic of most design techniques (including here the design leading to writing specifications) seems to be that they focus at one or at a few restricted aspects of the system operation which by themselves are not necessarily significant in the actual operating environment of the system. From the study of these aspects the designer extrapolates mentally the system performance to be expected within the actual operating environment. Guiding him in this extrapolation are his experience, engineering judgment, and some semiquantitative relations, the actual quantitative relationships frequently being too laborious to handle. In addition, the designer is also averaging mentally, on the basis of more or less conscious probability considerations regarding the utilization of the output with various inputs. In other words, while the synthesis is immediately concerned with various special facets of performance under largely hypothetical conditions the real concern is still the performance in the actual environment during the utilization of the output. Thus, looking behind all the indirectness of automatic feedback-control design, it develops that, although largely subconscious, actual system design is based on the one ultimate measure of performance: How much error is there at the times when the system output is utilized?

In this paper an attempt is made to give this measure of performance mathematical formulation which can be the basis of evaluating existing systems or the criterion for designing optimum systems. This measure is believed to be more general and to have more clear-cut physical meaning than existing criteria. In contrast to most standard techniques, which concentrate on isolated phases of performance such as transient or steady state, the proposed measure unites all of these, favoring none, in a single concept.

This paper is intended as the first of a group of papers. Accordingly, the basic philosophies are expounded here, the mathematical formulation and definition of the measure is introduced, and it is demonstrated that these definitions are amenable to mathematical analysis. General solutions rather than working equations are presented.

Proposed Measure of Performance Evaluation

An automatic feedback-control system exists to convert its input into a desired output which may or may not be identical to the input. The difference of desired and actual outputs is the error of the control system. Consequently the performance of the control system is characterized by the amount of the error at such times as the output is utilized. Clearly, this verbal statement of the evaluation of performance contains two elements: 1. the amount or size of the error; and 2. the times when the output is utilized. Both of these elements are subject, in their nature and relative significance, to the environmental conditions of the specific control system. For instance, in some control systems enlarged errors are increasingly, and possibly cumulatively, objectionable. Frequency control of some power systems might be an example. In other cases, such as in an automatic milling machine, errors up to a certain tolerance are not objectionable but all errors exceeding this tolerance are equally and totally objectionable. For most control systems (even for the examples mentioned) the lines are not so clearly drawn and the situation is intermediate between the two extremes. The point made is that the specific environment and function of

Paper 59-190, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 23, 1958; made available for printing December 16, 1958.

J. Zaborszky is with Washington University and consultant to McDonnell Aircraft Corporation, both St. Louis, Mo., and J. W. Diesel is with McDonnell Aircraft Corporation, St. Louis, Mo.

each control system will set a specific penalty valuation on errors of a given size. This penalty valuation can be expressed for the given application in the form of a penalty function $F(e)$, a single-valued function of the error e. In a more general sense, the penalty valuation may vary with extraneous parameters such as rate of flow in a temperature control system or with time, $F(e, t, v_1, v_2 \ldots v_r)$, where v_1 through v_r are the parameters. Of course error, e, itself certainly is a function of time, t.

The second element of performance evaluation, the time when the output is utilized, shows equally wide individual variation depending on the environment of the control system. Some control systems, such as a frequency or voltage control on a power system, operate at all times on essentially steady reference inputs and a stationary random noise input, the varying load. All times for utilizing the output are equally probable here. The other extreme would be a control system which makes one adjustment upon activation by a step input followed by one instantaneous measurement after a fixed-time interval. Here the time of utilization is one instant after a fixed-time interval which has probability one while all other time intervals have probability zero. If the measurement itself takes a certain time, all times within this measuring time are equally probable. In a missile, the output is utilized at one single instant, but over a number of launchings the time intervals elapsing from launching to explosion will form a probability distribution centering around the nominal range.

Essentially, then, for any automatic feedback-control system, the times elapsing from activating the system to all times of utilization of its output can be arranged into a probability distribution, $p(t)$, over many repeated activations by identical inputs of the same system or identical systems.

Mathematical Formulation of the Measure of Performance

The measure of performance described in the preceding section can be mathematically formulated as

$$\mathfrak{s} = \int_0^\infty \overline{F(e(t), t, v_1 \ldots v_r)} p(t) dt \qquad (1A)$$

where the symbol $\mathfrak{s}$ is the end sigma, the form of letter sigma used at the end of Greek words.

For one specific type of input, a "deterministic" input, equation 1(A) clearly gives the statistically correct average value of the penalized error at all times when the output is utilized.

If the environment of the control system includes a number of different types of inputs, possibly an infinite number or ensemble of different types of inputs, an averaging over, this ensemble will also be necessary to obtain the mean penalized error. Then the measure of performance takes the form

$$\mathfrak{s} = \int_0^\infty \overline{F(e(t), t, v_1 \ldots v_r)} \overline{p(t)} dt \qquad (1B)$$

where the straight line denotes the averaging over the ensemble, so that equation 1(B) describes the average penalized error.

Frequently the probability distribution of utilization $p(t)$ will be independent of the type of the input. Then equation 1(B) reduces to

$$\mathfrak{s} = \int_0^\infty \overline{F(e(t), t, v_1 \ldots v_r)} p(t) dt \qquad (1C)$$

The value of the proposed measure of performance as defined in equations 1(A)–1(C) is conceptually the average value of penalized error at such times when the output is utilized, and for purposes of identification it will be referred to by either of the names "probabilistic error" or "end sigma error." As a performance measure or criterion equations 1(A)–1(C) will be referred to as the "end sigma criterion."

This measure unites in a single concept the transient and steady states of the system operation as well as the largely neglected intermediate states. None of these operating states is discriminated for or against and because of $p(t)$ each gets the weight due it because of the relative frequency of its presence at such times as the output is utilized.

In addition to uniting the various states of system performance, equation 1 also unites most of the various criteria used to evaluate these states separately. As an example, if $F(e, t, v_1 \ldots v_n) = e^2$, then referring to equation 1(B)

$$\mathfrak{s} = \int_0^\infty \overline{e^2(t) p(t)}\, dt \qquad (2)$$

becomes the true mean-square value of the error. This end sigma, or probabilistic error, unifies in one concept the mean square criterion used previously for stationary[1,2] and nonstationary[3–4] random input environments and the integral square criterion used for nonstationary random[3] and deterministic[3] environments. These become special cases of equation 2. Final value systems[4] also are special cases of equation 1 by the use of a probability function as given in equation 11. However, equation 2 is not restricted to these special cases. This is general in that it truly averages the square error whatever may be the operating state of the system.

Performance criteria similar in form to equation 1, but basically different in meaning, have been proposed.[7–14] The basic difference is that instead of the probability distribution function, $p(t)$, of equation 1, an arbitrary weighting function is proposed either in a concrete[8–14] or a general form.[13,14] Consequently these criteria define arbitrarily weighted penalty factors without clear physical meaning. Nevertheless, equations 1(A)–1(C) include as special cases these arbitrarily weighted penalty criteria. Assume for instance

$$F(e(t), t, v_1 \ldots v_r) = F(e) W(t) \qquad (3)$$

where $W(t)$ is a function weighting, or penalizing, the errors according to the time when they occur. $W(t) = t$ and $W(t) = t^2$ have been proposed and investigated.[8–14] Then, for instance, equation 1(C) reduces to

$$\mathfrak{s} = \int_0^\infty \overline{F(e)} g(t) dt \qquad (4)$$

where

$$g(t) = W(t) p(t) \qquad (5)$$

now becomes a modified probability function which, in addition to the relative frequency of occurrence of various utilization times, also includes the designers opinion regarding the relative importance of these utilization times. Equations 1(C) and 4, being identical in form, can be handled by identical techniques which are introduced in subsequent parts of this paper. It is significant that these techniques make it possible to obtain mathematical solutions for the optimum systems under the earlier performance criteria, something which has been done heretofore only by cut and try.

Using equation 1, it is not necessary to define a number of separate performance criteria. Transient and steady conditions, deterministic or random input environments, linear and nonlinear systems are all covered by the same criterion.

Special Cases of Penalty Functions

If the magnitude of the error alone is to be the measure of its undesirability, then

$$F(e) = |e| \qquad (6)$$

is the penalty function. The widely used penalty function

$$F(e)=e^2 \tag{7}$$

of course puts an increasingly high penalty on increasing errors. This property does not fit such cases as automatic machine tools or missiles where all errors up to a certain tolerance are almost equally acceptable and all errors beyond the tolerance are equally unacceptable. For such control systems

$$F(e)=\frac{e^2}{k+e^2} \tag{8}$$

with k a constant, would be quite suitable since this penalty function (Fig. 1) penalizes errors progressively up to a certain value and then puts an essentially constant penalty on errors past this value.

Even more suitable would be the following:

$$F(e(t))=1-\int_{-e_t}^{e_t} p(e|t)de=p_t(t) \tag{9}$$

which gives the probability of error, e, being outside the tolerance value, e_t at time t. In equation 9, $p(e|t)$ is the conditional probability distribution in e and t. For the penalty function of equation 9, equation 1 yields

$$\epsilon=\int_0^\infty p_t(t)p(t)dt=1-\int_{-e_t}^{e_t}\int_0^\infty p(e|t)p(t)dedt$$

$$=1-\int_{-e_t}^{e_t}\int_0^\infty p(e,t)dedt \tag{10}$$

which is simply the over-all probability of the error being outside the tolerance, a most appropriate criterion, which in the case of a missile could readily be extended to the over-all hit probability.

It should be emphasized that the specific $F(e)$ function is determined by the nature of the problem and is to be established by the same engineering judgment which establishes countless other factors of the design. The penalty function $F(e)=e^2$ is the most commonly used, and it will be used exclusively in the remainder of the paper.

Special Cases of Probability Distribution Functions

By definition, $p(t)$ is the probability distribution of all times when the output is being utilized, based on a large number of repeated applications of inputs of one given type. The starting point of the time scale $t=0$ for determining $p(t)$ is the instant of activation

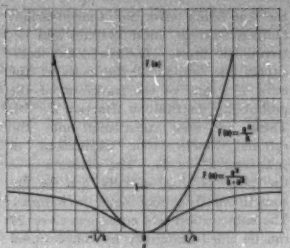


Fig. 1. Two penalty functions

of the control system. For a missile, $t=0$ can signify the instant of firing or the instant of transition to the homing phase. For some other system, $t=0$ may be the instant of turning on the power in the control system or of the arrival of a step input, etc.

The probability distribution function, $p(t)$, can be readily established in some cases. In others, exact determination would require an extensive operations study, but a reasonable estimate can almost always be found quite readily. As a matter of fact, by a slight shift of the sense of the definition, $p(t)$ could be represented as expressing the distribution to which the spread of the possible utilization times can be narrowed down on the basis of the limited information available to the designer. In this sense, $p(t)$ would become, at least partly, representative of the lack of definition in the design parameters rather than a probability distribution of the actual operating times.

As an example, consider a control system which receives step inputs at $t=0$ followed by one instantaneous measurement at $t=T_0$.

Then

$$p(t)=\delta(t-T_0) \tag{11}$$

where $\delta(t)$ denotes the Dirac delta function. If in the same system the measurement extends from T_0 to T_1 (as it would for instance in some wind-tunnel applications)

$$p(t)=\frac{1}{T_1-T_0}\,[u(t-T_0)-u(t-T_1)] \tag{12}$$

where $u(t)$ denotes the unit-step function.

In writing equations 11 and 12 the assumption was that the times of utilization are the same for every repeated application of the step input. Should these be variable in some way (in the case of a missile the instant of explosion, T_0 in equation 11, would vary from launching to launching depending on the distance of the target at the time of launching) then equations 11 and 12 modify respectively into

$$p(t)=\lim_{K=\infty}\frac{1}{K}\sum_{k=1}^{K}\delta(t-T_{0k})=\overline{\delta(t-T_0)} \tag{13}$$

$$p(t)=\lim_{K=\infty}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_{1k}-T_{0k}}\times[u(t-T_{0k})-u(t-T_{1k})]$$

$$=\overline{\frac{1}{T_1-T_0}\,[u(t-T_0)-u(t-T_1)]} \tag{14}$$

For a missile where maximum and minimum ranges (T_1 and T_0 respectively) are clearly set, the results of equation 13 would come close to the form of equation 12.

For systems which operate essentially in the steady state, one possible form can be produced from equation 12:

$$p(t)=\lim_{T_1=\infty}\frac{1}{T_1}\,[u(t)-u(t-T_1)] \tag{15}$$

or from a Rayleigh probability distribution:

$$p(t)\begin{cases}=\lim_{\sigma=0}\dfrac{t}{\sigma^2}\exp\left(-\dfrac{t}{2\sigma^2}\right) & t\geq 0\\ =0 & t<0\end{cases} \tag{16}$$

With either equation 15 or 16, equation 2 goes over into the customary definition of the rms error as used in the study of systems subjected to stationary random processes. This is especially obvious in connection with equation 15.

Finding the Probabilistic Error for Linear Systems in s Domain

The end-sigma mean-square error is defined by equation 2. Its computation for linear-time-invariant lumped-constant systems can be carried out either in the s domain of the Laplace transforms or in the time domain. The s-domain approach will be discussed first.

The Laplace transform of the error can be expressed as

$$E(s)=I(s)-K(s)R(s)=\frac{M(s)}{N(s)} \tag{17}$$

where $M(s)$ and $N(s)$ are entire functions with $N(s)$ having a higher order singularity than $M(s)$. $I(s)$ and $R(s)$ are, respectively, the Laplace transforms of the desired output and the actual input, and $K(s)$ is the closed-loop transfer function of the system. Now, as shown in the Appendix, equation 2 for a deterministic input $[e^2(t)=e^2(t)]$ can be converted by complex convolution into

$$\epsilon = \lim_{t=0} \mathcal{L}_2^{-1}\{P(-s)[E(s) * E(s)]\} \quad (18)$$

where $\mathcal{L}_2^{-1}$ denotes the inverse 2-sided Laplace transform.[15] $P(s)$ is the 1-sided Laplace transform of $p(t)$, and the asterisk denotes complex convolution. Since $p(t)$ is zero for negative time ($t=0$ being the instant of activation, so that any utilization of the output can come only in positive time),

$$P(-s) = \mathcal{L}_2[p(-t)] = \frac{A(s)}{B(s)} = \int_{-\infty}^{\infty} p(t)e^{st}dt \quad (19)$$

where $A(s)$ and $B(s)$ are entire functions. Further,

$$E(s) * E(s) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} E(s-w)E(w)dw = \sum_i \frac{M(s-p_i)M(p_i)}{N(s-p_i)N'(p_i)} = \frac{J(s)}{L(s)} \quad (20)$$

where $J(s)$ and $L(s)$ are entire functions, the p_i are the zeros of $N(s)$, all of which are assumed to be single, and

$$N'(s) = \frac{d}{ds}N(s) \quad (21)$$

By using inverse transformation through contour integration, equation 18 goes over into one of the two alternate forms of equation 22 and equation 24

$$\epsilon = -\sum_k \frac{A(q_k)}{B'(q_k)}\frac{J(q_k)}{L(q_k)} \quad (22)$$

where the q_k are the poles of $P(-s)$, all of which are assumed to be single, and all are located in the right half plane. Also

$$B'(s) = \frac{d}{ds}B(s) \quad (23)$$

Alternatively,

$$\epsilon = \sum_i \frac{A(p_i)}{B(p_i)}\frac{J(p_i)}{L'(p_i)} \quad (24)$$

where the p_i are the zeros of $L(s)$, all of which are assumed single, and located in the left half plane. Also

$$L'(s) = \frac{d}{ds}L(s) \quad (25)$$

Two equations (equations 22 and 24) result from the 2-sided inverse transformation since the function is continuous at $t=0$ and, consequently, its limit can be evaluated either from the transform for positive time or from the transform for negative time, in other words, by closing the contour in the s plane on either the right or left.

Equations 22 and 24 represent the general s-plane formulas for evaluating the end-sigma error for time-invariant lumped-constant linear systems and deterministic inputs. The evaluation necessitates the finding of the poles of the error $E(s)$ and of the transform of the probability distribution function $P(-s)$. These are then substituted into the rational function-type expressions of equations 22 or 24. The evaluation of the poles of $E(s)$ usually can be accomplished by root-locus techniques. The substitution into equations 22 or 24 can be aided materially by conventional spirule operations. Inherent in these derivations is the assumption that $E(s)$ has no pole at the origin.

For example, assume an input

$$r(t) = i(t) = 1+2t \qquad t>0$$

and probability function

$$p(t) = \frac{10}{3}(e^{-3t} - e^{-5t})$$

while the system closed-loop transfer function is

$$K(s) = \frac{s+1}{s^2+s+1}$$

Then

$$R(s) = I(s) = \frac{s+2}{s^2}$$

$$E(s) = I(s) - K(s)R(s) = \frac{s+2}{s^2+s+1}$$

so that $p_1 = -0.5+j\sqrt{0.75}$ $p_2 = -0.5-j\sqrt{0.75}$

Also

$$P(-s) = \frac{10}{(s-2)(s-5)}$$

so that $q_1 = +2$, $q_2 = +5$.

From equation 20

$$\frac{J(s)}{L(s)} = \frac{(s-2p_2)(s-p_1+2)(p_1+2) - (s-2p_1)(s-p_2+2)(p_2+2)}{(s-2p_1)(s-2p_2)(s-p_1-p_2)(p_1-p_2)}$$

and the end-sigma error from equation 22 is $\epsilon = 1.37$.

This highly simplified example should be interpreted as illustrative of the computation process and not as illustrative of the power of the method attainable when working equations are developed from general solutions given in this paper.

Finding the Value of the Probabilistic Error in the t Domain

If the end-sigma error defined by equation 2 in the t domain is to be evaluated in the time domain, first it will be assumed that the system is time-invariant linear with known impulsive response or weighting function, $k(t)$. Consequently the error can be defined by time-domain convolution, or superposition, as

$$e(t) = i(t) - \int_0^{\infty} k(t_1)r(t-t_1)dt_1 \quad (26)$$

where $i(t)$ is the desired output, $r(t)$ is the input, and $k(t)=0$ for $t<0$ which is necessary for a realizable $k(t)$.

Let the following function be defined

$$\sigma^2(t,\tau) = \int_0^{\infty} p(t_2+t)e(t_2)e(t_2+\tau)dt_2 \quad (27)$$

Then

$$\epsilon = \sigma^2(0,0) \quad (28)$$

Substituting equation 26 into equation 27, and rearranging by an exchange of the sequence of integration and considering equation 28, there results

$$\epsilon = \psi_{ii}(0,0) - 2\int_0^{\infty} dt_1 k(t_1)\psi_{ii}(0,-t_1) + \int_0^{\infty} dt_2 k(t_1)\int_0^{\infty} dt_1 k(t_1)\psi_{rr}(t_2,t_2-t_1) \quad (29)$$

where the following definitions were made:

$$\psi_{rr}(t,\tau) = \int_0^{\infty} dt_2 \overline{p(t_2+t)r(t_2)r(t_2+\tau)} \quad (30)$$

$$\psi_{ii}(t,\tau) = \int_0^{\infty} dt_2 \overline{p(t_2+t)i(t_2)i(t_2+\tau)} \quad (31)$$

$$\psi_{ir}(t,\tau) = \int_0^{\infty} dt_2 \overline{p(t_2+t)i(t_2)r(t_2+\tau)} \quad (32)$$

where the straight line denotes ensemble averaging for the conditions of equation 1(B).

Equation 29, then, is the time-domain equivalent of equations 22 or 24. The evaluation of the integrals in equation 29 yields the end-sigma error value for a given system. The results are readily generalized to cases of finite memory transfer functions, ($k(t)=0$ if $t>T$), or probability density functions of finite duration, ($p(t)=0$ if $t>T$), by simply changing the upper limit of the integrals to T in equations 1 and 27–32.

Some additional discussion, however, is due equations 30–32. These equations define a generalized form of auto- and cross-correlation functions which unifies in one concept the auto- and cross-correlation functions customarily used for stationary processes[1] and the auto- and cross-translation functions proposed for transient conditions.[2] Clearly the probability function of equation 15

immediately reduces equations 30 through 32 to the definitions of auto- and cross-correlation functions. The same probability functions would result in a modified but equivalent definition of the translation functions.[3] An infinitude of such special cases result from the various particular probability functions.

Taking a time varying Laplace transform[4] of equations 30 through 32, a generalized power-spectral density corresponding to the general correlation functions would result.

Referring to equation 30,

$$\Psi_{rr}(t,s)=\int_{-\infty}^{\infty}\psi_{rr}(t,\tau)e^{s\tau}d\tau$$

$$=\int_{0}^{\infty}p(t_1+t)\Phi(t_1,s)dt_1 \qquad (33)$$

where

$$\Phi(t_1,s)=\int_{-\infty}^{\infty}\overline{r(t_1)r(t_1+\tau)}e^{-s\tau}d\tau \qquad (34)$$

is the apparent power-spectral density for the ensemble at time t_1. If the ensemble is stationary and the ergodic theory applies,

$$\Phi(t_1,s)=\Phi(s) \qquad (35)$$

and the generalized power spectral density,

$$\Psi(t_1,s)=\Phi(s)\int_{-\infty}^{\infty}p(t_1+t)dt_1=\Phi(s) \qquad (36)$$

reduces to the conventional power-spectral density of the stationary process, as it of course should. Similar definitions are easily obtained from equation 32 for a generalized cross-power spectral density.

Finding the Optimum Invariant Linear System Under the End-Sigma Criterion

Equations 22, 24, and 29 are useful in evaluating the end-sigma error for a system of known construction and characteristics by working respectively in the s or t domains. If a choice exists between several potential systems, finding and comparing $\mathfrak{s}$ for each of these immediately will pinpoint that particular system which gives optimum error performance, that is, minimum true rms error. This would be an optimum from among a limited choice, established by cut-and-try techniques. Now if the choice of systems is made unlimited to the point of including all linear systems with constant coefficients, then a direct solution for the optimum system becomes possible, by finding the optimum $k(t)$ weighting function or its transform, the $K(s)$ transfer function. Since either of these unknown functions is included in the defining integral of $\mathfrak{s}$, the quantity to be minimized (equations 18 and 29 respectively), the solution can be obtained by calculus of variations. This will be illustrated here for the time domain, equation 29. The minimum of $\mathfrak{s}$ is defined by equating its first variation, $\delta\mathfrak{s}$, to zero, that is from equation 29

$$\delta\mathfrak{s}=2\int_{0}^{\infty}d\delta k(t)\times$$

$$\left[\int_{0}^{\infty}dt_1k(t_1)\psi_{rr}(t,t-t_1)-\psi_{ir}(0,-t)\right]=0 \qquad (37)$$

which is only possible for arbitrary $\delta k(t)$ if

$$\int_{0}^{\infty}dt_1k(t_1)\psi_{rr}(t,t-t_1)-\psi_{ir}(0,-t)=0 \quad t>0 \qquad (38)$$

It can be shown that equation 38 actually defines a minimum.

As in the case of equation 29, equation 38 is also easily adapted to finite memory-transfer functions and finite duration probability distribution.

Equation 38 is a linear integral equation of the first kind. $\psi_{rr}(t,t-t_1)$ is its kernel and as may readily be shown from equation 30 this kernel is symmetric. The solution of this integral equation for $k(t)$ is the impulsive response of the optimum system, the system yielding the minimum end-sigma error. The latter then should be found from equation 29.

The literature, both mathematical[10] and engineering,[2] for solving integral equations is extensive. No attempt toward completeness is made in the references.

Conclusions

As a measure of performance, the end sigma or probabilistic error eliminates the indirectness so characteristic of usual design techniques. The end-sigma error summarizes the entire system performance, rather than explaining it in full detail, thus making it a desirable starting point for analysis or design. Such techniques as s-plane transient analysis can then be used to spotlight special limited aspects of the performance.

In addition to introducing this measure of performance it has been demonstrated that it can be evaluated in either the s or t domain for time-invariant linear systems. An integral equation of a common type has been introduced to define the impulsive response of the optimum-invariant linear system under the end-sigma criterion.

Appendix

Equation 2 can be rephrased as

$$\mathfrak{s}=\lim_{t=0}\int_{0}^{\infty}p(\tau-t)e^{s}(\tau)d\tau \qquad (39)$$

Or introducing 2-sided Laplace transformation[10,17] $\mathcal{L}_2[\ ]$ and inverse 2-sided Laplace transformation $\mathcal{L}_2^{-1}[\ ]$

$$\mathfrak{s}=\lim_{t=0}\mathcal{L}_2^{-1}\left[\mathcal{L}_2\left[\int_{0}^{\infty}p(\tau-t)e^{s}(\tau)d\tau\right]\right]$$

$$=\lim_{t=0}\mathcal{L}_2^{-1}\left[\int_{-\infty}^{\infty}e^{-st}dt\times\right.$$

$$\left.\int_{0}^{\infty}p(\tau-t)e^{s}(\tau)d\tau\right]$$

$$=\lim_{t=0}\mathcal{L}_2^{-1}\left[\int_{0}^{\infty}e^{s}(\tau)e^{-s\tau}d\tau\times\right.$$

$$\left.\int_{-\infty}^{\infty}p(\tau-t)e^{s(\tau-t)}d(t-\tau)\right] \qquad (40)$$

Further, from complex convolution theory

$$\mathcal{L}_2[e^{s}(\tau)]\int_{0}^{\infty}e^{s}(\tau)e^{-s\tau}d\tau=E(s)*E(s) \qquad (41)$$

Equations 40, 41, and 19 can be combined into equation 18.

References

1. Extrapolation, Interpolation, and Smoothing of Stationary Time Series (book), N. Wiener. John Wiley & Sons, Inc., New York, N. Y., 1949.

2. Analytical Design of Linear Feedback Controls (book), G. C. Newton, Jr., L. A. Gould, J. F. Kaiser. John Wiley & Sons, Inc., 1957.

3. Random Processes in Automatic Control (book), J. H. Laning, Jr., R. H. Battin. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

4. An Extension of Wiener's Theory of Prediction, L. A. Zadeh, J. R. Ragazzini. *Journal of Applied Physics*, New York, N. Y., vol. 21, 1950, p. 645.

5. An Optimization Theory for Time-Varying Linear Systems with Non-stationary Statistical Inputs, R. C. Booton, Jr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Aug. 1952, p. 977.

6. Final Value Controller Synthesis, M. V. Mathews, C. W. Steeg. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-2, Feb. 1957, p. 6.

7. A Note on Control Area, T. M. Stout. *Journal of Applied Physics*, New York, N. Y., vol. 21, 1950, p. 1120.

8. Some Design Criteria for Automatic Controls, Paul T. Nims. *AIEE Transactions*, vol. 70, pt. 1, 1951, pp. 606–11.

9. The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms, D. Graham, R. C. Lathrop. *Ibid.*, vol. 72, pt. II, *Applications and Industry*, Nov. 1953, pp. 273–88.

10. The Transient Performance of Servomechanisms with Derivative and Integral Control, Richard C. Lathrop, Dunstan Graham. *Ibid.*, vol. 73, pt. II, Mar. 1954, pp. 10–17.

11. The Influence of Time Scale and Gain on Criteria for Servomechanism Performance, Dunstan Graham, Richard C. Lathrop. *Ibid.*, July, pp. 153–58.

12. An Integral Criterion for Optimizing Duplicator Systems on the Basis of Transient Response, Jack Crow. *Doctoral Dissertation*, Washington University, St. Louis, Mo., 1956

13. The Selection and Use of Servo Performance Criteria, W. C. Schultz, V. C. Rideout. *AIEE Transactions*, vol. 76, pt. II, 1957 (Jan. 1958 section), pp. 383–88.

14. A General Criterion for Servo Performance, W. C. Schultz, V. C. Rideout. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 13, 1957, pp. 549–60.

15. Theorie und Anwendung der Laplace-Tranformation (book), G. Doetsch. Dover Publications, Inc., New York, N. Y., 1943.

16. Linear Integral Equations (book), W. V. Lovitt. Dover Publications, Inc., 1950.

17. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., 1955.

Modern Systems of Traffic Control as Applied to the Seaboard Air Line Railroad Company

J. R. DePRIEST
NONMEMBER AIEE

ON December 6, 1941, the Seaboard Air Line Railroad placed in service the first segment of a major installation of a system of traffic control on its line. The segment was only 33 miles in length but it made history. It was a portion of a 65-mile segment of railroad to be electrically controlled from a remote location 157 miles from the most distant point. At that time this line became the longest control line for this type of service in the world. The system proved to be very efficient and economical, and expansion of the system was made as rapidly as material and labor could be obtained. As of today, trains operate the entire distance between Richmond and Miami–Tampa by signal indication controlled by modern systems of traffic control.

Prior to December 6, 1941, the Seaboard line consisted of a single-track railroad with more than 4,000 miles of track extending from Richmond to Miami and Tampa. A main line also extended from Hamlet, N. C., toward Atlanta and Birmingham together with other extensions operating in the states of Virginia, North Carolina, South Carolina, Georgia, Alabama, and Florida. At that time the Seaboard had approximately 550 miles of automatic signals mainly extending between Richmond, Va., and Monroe, N. C.; between Savannah, Ga., and Baldwin, Fla.; and between Starke and Coleman, Fla.; see Fig. 1. With the systems then in service it was necessary for telegraph operators or telephone operators to be placed at strategic points along the railroad in order to receive train orders issued by a train dispatcher. The operator then physically delivered these orders to train crews in order that opposing trains could meet at strategically placed sidings at the proper time. Since it takes considerable time to issue and check train orders, it can be seen that the system becomes difficult and expensive to operate as the number of trains increases. This is so because the total number of meets in a fixed distance increases directly in proportion to the square of the number of trains. Perhaps, this point should be further discussed so that the reader may become familiar with the problems encountered in designing a traffic-control system for operating a single-track railroad.

It is evident that for a double-track railroad it is possible to operate trains of equal speeds at equal time intervals without interference from either following or opposing trains, since trains moving in opposite direction operate on adjacent tracks. Fig. 2(A) is a time-distance chart showing the position of southward (*1S*) and northward (*1N*) trains. Train *1S* leaves station *A* proceeding toward station *C* at a constant rate of speed. The progress of this train is represented by a diagonal line with a negative slope. At the same time train *1N*, represented by a diagonal line with a positive slope, operating in the opposite direction, leaves station *C* proceeding toward station *A* at the same speed. Since each train is operating at the same constant speed they will arrive at their destinations *C* and *A*, respectively, after a time interval t, having traversed the distance D between stations. It will be noted from Fig. 2(A) that the two diagonal lines, representing the progress of train *1S* and *1N*, intersect or cross one another half way between stations *A* and *C*, after one half of the time interval t has elapsed. Since train *1S* leaving station *A* will meet only one train between stations *A* and *C* (and one train at *C* if the railroad continues for a greater length) the railroad could be single track with the exception of a small distance near the center (siding *B*) and at stations *A* and *C* if means were provided whereby one of the trains could enter, pass the opposing train, and then leave the siding. It should be pointed out that if train *1S* continued beyond station *C* another train would be met at station *C* if opposing trains were dispatched at equal intervals.

Similarly Fig. 2(B) illustrates the conditions encountered if the number of trains operated in time interval t were doubled. It can be seen from Fig. 2(B) that the total number of meets occurring within a time interval t would increase from two to eight. However, four sidings would be required as compared with two sidings previously. If 3 trains in each direction were operated within time interval t, then by referring to Fig. 2(C) it can be seen that a total of 18 meets would occur within time interval t and 6 sidings would be needed.

It is interesting to note that where one train in each direction is operating per time interval t, siding *B* is used one time as shown by the heavy dot located vertically under siding *B*. Where two trains in each direction are operating per time interval t, siding *B* is used twice. And where three trains are operating in each direction per time interval t, three trains use siding *B*.

If t is made equal to a time interval of one day and Table I is prepared from the foregoing information, a formula can be written from the first three conditions to determine the total number of

Paper 59-251, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 16, 1958.

J. R. DePriest is with the Seaboard Air Line Railroad Company, Richmond, Va.

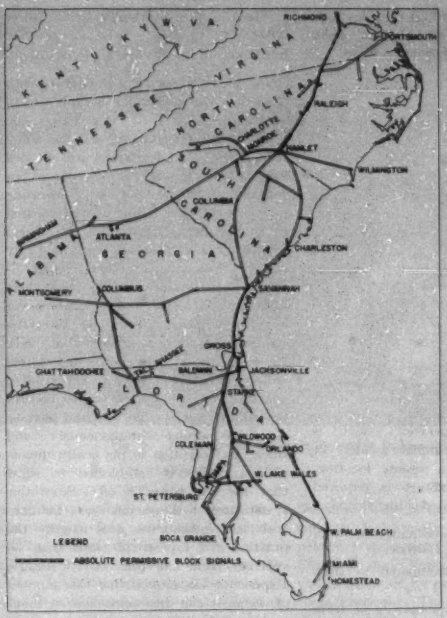


Fig. 1 (left). Type of signals on Seaboard prior to Dec. 6, 1941

Table I

T	t	M	S
1	1	2	2
2	1	8	4
3	1	18	6
4	1	32	8
5	1	50	10
10	1	200	20
20	1	800	40
24	1	1,158	48
48	1	4,608	96
72	1	10,368	144
96	1	18,432	192

meets occurring within the area the train operates in one day or for any time interval selected. Similarly, formulas can be derived showing the meets per train per day which equals the number of sidings needed within the area. These formulas are as follows:

$$M = 2(Tt)^2$$

$$S = M/Tt$$

$$S = 2Tt$$

where

t = time interval a train operates in traveling distance D

T = number of trains operated per direction per time interval t

M = total number of meets occurring within time interval t or in distance D

D = distance train travels in time interval t

S = number of sidings needed within distance D or time interval t

For a practical example, assume t is equal to one day and one train is operated each half-hour, then 48 trains will be operated in each direction per day. From Table I, under S, 96 sidings will be needed. If the trains average 40 mph, a train will travel 960 miles in one day, which is distance D. Therefore, if the 960 miles is divided by the 96 sidings it will be found that sidings must be spaced 10 miles apart. It should be pointed out that with sidings spaced 10 miles apart and trains averaging 40 mph, if one of the trains was delayed and could not make the scheduled progress, it would be necessary to require the delayed train to enter the siding one station before its scheduled point to avoid delay to the opposing train. This train would then be further delayed for a time interval not exceeding the sum of its running time to the next station and the running time of the opposing train from the next station to the siding where the meet will occur. Where the speeds of each train are equal, then the maximum additional delay that could occur would be two times the running time between sidings, or in this case 18 minutes. To reduce this type of delay more sidings are usually provided than required to handle the maximum number of trains as shown by Table I.

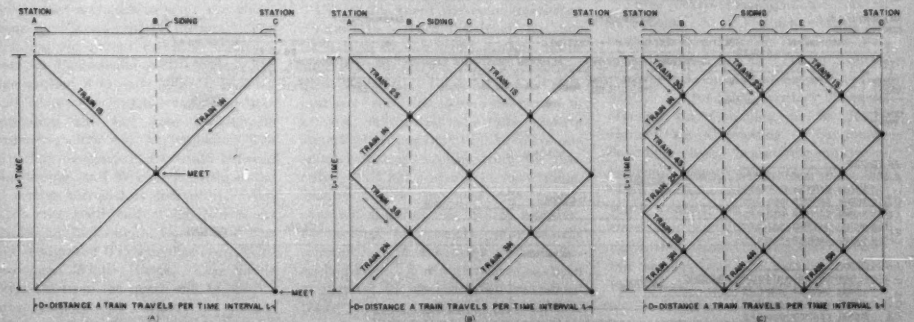


Fig. 2. Time-distance chart

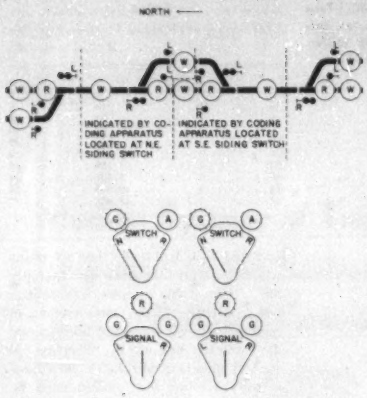


Fig. 3. Track diagram with associated indication lights and control switches

A traffic-control system remotely controlled must provide electric power to operate and control a railroad switch located at the entrance or exit of a siding, and extend information to enginemen of trains by means of wayside signals appropriately located. It is necessary for the dispatcher to control the following functions for each siding switch:

1. Position switch normal.
2. Position switch reverse.
3. Position northward signals to either stop or clear.
4. Position southward signals to either stop or clear.
5. Position automatic reclearing feature for following trains.
6. Operate maintainer's call light.

The system must indicate certain information from the field location, where the power-operated switch is located, to the dispatcher located at control machine at a remote location. The following information is required from each power-operated siding switch:

1. Occupancy or nonoccupancy by train located on track north of switch.
2. Occupancy or nonoccupancy by train located on switch.
3. Occupancy or nonoccupancy of train located south of switch either on main line or siding. (Usually a typical procedure is to assign siding occupancy indications to apparatus located at north switch of siding and occupancy of track between switches to apparatus located at south switch.)
4. Indication to show when switch is set for main-track traffic or in transit position.
5. Indication to show when switch is set for side-track traffic or in transit position.
6. Indication of northward signal aspect (clear or stop).
7. Indication of southward signal aspect (clear or stop).
8. Power off indication (on or off).

Fig. 3 illustrates a track diagram with associated indication lights and control switches.

From the foregoing it will be noted that 6 bits of information are needed for control purposes and 8 bits of information are needed for indicating purposes at each siding switch located in traffic-control territory. The Union Switch *506-A* system, described by Mr. Baughman[1] in another paper, is arranged to control and indicate by code 7 bits of information for 35 stations over one pair of wires. From the foregoing requirements it can be seen that, with the exception of indications, this system will be adequate. Since power is off a small percentage of time, it is practical to indicate both a normal and reverse switch condition simultaneously, and since this condition is physically impossible a means is established to show that the a-c power is off. Since this system uses a share-time basis for controls and indications, it is usually the practice not to control more than 30 switches from one line, or 15 sidings. Experience has proved that this number of sidings can be controlled without unduly delaying indication or control codes for all stations on a system used by a busy railroad. If more than 15 sidings are needed, the territory is divided in two or more sections and a system is applied to each section. When this is done, a pair of wires extending from the most distant siding switch to the control point is necessary for each section of line to which a system is applied. If the d-c coded impulses at the end of each section are used to

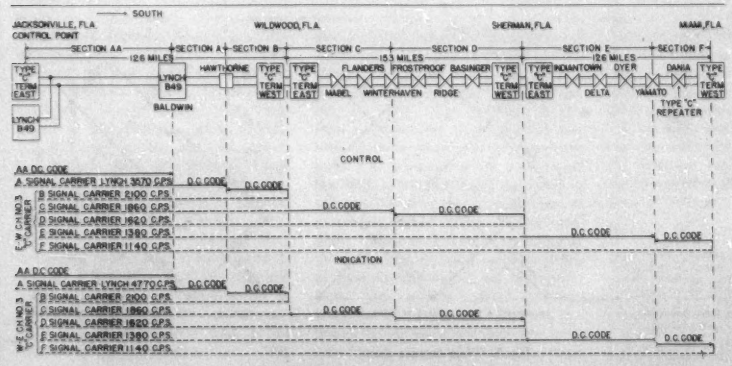


Fig. 4. Diagram of carrier sections

Fig. 5. Traffic control machines located in Jacksonville

modulate a voice-frequency telegraph-type carrier, only one pair of wires will be needed throughout the entire railroad for this purpose. For control codes, similar carrier apparatus is used but the modulation takes place at the control machine and the carrier is demodulated at an end of the remote section in the field. Fig. 4 illustrates the control arrangement used by the Seaboard Air Line Railroad between Jacksonville and Miami, Fla.

It will be noted in the foregoing arrangement that a Western Electric *C* carrier which provides three trunk telephone circuits was applied to the single pair of wires of the code line and cable extending between Jacksonville and Wildwood, Wildwood and Sherman, and Sherman to Miami. One channel of this carrier was used to apply the voice-frequency signal-control carriers. The first line or system called *AA* extends between Jacksonville and Baldwin. The line marked *A* extends between Baldwin and Hawthorne with carrier controls extending this line into Jacksonville from its north end at Baldwin. The next line designated *B* extends between Hawthorne and Wildwood. Since the *C* carrier terminal is located at Wildwood, the south end of this line is extended to Jacksonville by carrier. The line designated *C* extends between Wildwood and Winter Haven. This line is extended from the north end by carrier into Jacksonville. Line *D* extends between Winter Haven and Sherman and since carrier *C* terminal is located at Sherman, it is extended from the south end to Jacksonville by carrier. The next line designated *E* extends between Sherman and Yamato and the north end is extended by carrier from Sherman to Jacksonville. The last line designated *F* extends between Yamato and Miami and since the *C* carrier terminal is located at Miami the *F* is extended from its south end to Jacksonville. The distance from Jacksonville to Miami is approximately 405 miles. The controls for this entire territory are placed on three machines, one controlling the territory between Jacksonville and Baldwin, another between Baldwin and Wildwood, and the other controlling the territory between Wildwood and Miami. Fig. 5 shows the traffic control machines located in Jacksonville.

Up to this point the communications involved have been discussed, but not the actual wayside signal systems available for traffic control. On the Seaboard three distinct types of systems have been installed. The first system is referred to as the full-centralized traffic control system which permits trains to follow one another without waiting until the preceding train has reached the next siding. This system is equipped with power-operated switches and, where required, the sidings are usually more closely spaced than in the other systems described later. The next system has been designated as the modified-centralized traffic control system. This system provides for power-operated switches, but will not permit trains to follow until the preceding train has reached and passed the entrance switch to the next siding. The system eliminates the many wayside signals needed when following movements are permitted as in the full-centralized traffic control. The third type of centralized traffic control is designated by the Seaboard as remote-controlled automatic signals system. This system basically is arranged for train operation by signal indication in the same manner as the modified-centralized traffic-control system, except that following trains cannot follow another train until the first train has reached and passed the north end of the siding ahead. It can be seen that this system has eliminated much apparatus at one end of each siding. The siding switches are usually hand operated. However, the system is so designed that power-operated switches can be added when needed at a minimum of expense and the system can be expanded into either modified-centralized traffic control or full-centralized traffic control by adding the necessary apparatus for such system. In addition to the traffic control systems the Seaboard has short stretches of double track where the conventional automatic signals are provided and where trains operate by signal indication with current of traffic. Fig. 6 shows the various types of centralized traffic control used on the Seaboard as of October 1958. In addition the map shows the absolute-permissive block signal system between Hamlet and Atlanta. This system is a conventional automatic-permissive block signal system and is used with train orders for the movement of trains. It will be noted from this map that the Seaboard operates by signal indication for the entire distance of more than 1,040 miles between Richmond, Miami, and Tampa.

Three control machines are located at Raleigh, N. C., which control the entire territory between Richmond, Va., and Hamlet, N. C. Three control machines are located at Savannah, Ga., which control the territory between Hamlet via Columbia to Savannah; Hamlet via Charleston to Savannah; Savannah–Gross–Jacksonville and Gross–Baldwin, respectively. At Jacksonville there are three traffic control machines, which control the entire territory between Jacksonville and Miami. One machine controls the Jacksonville-Baldwin section, one the section between Baldwin and Wildwood and the other the section between Wildwood and Miami. The section between Coleman and Tampa, Florida, is controlled from a

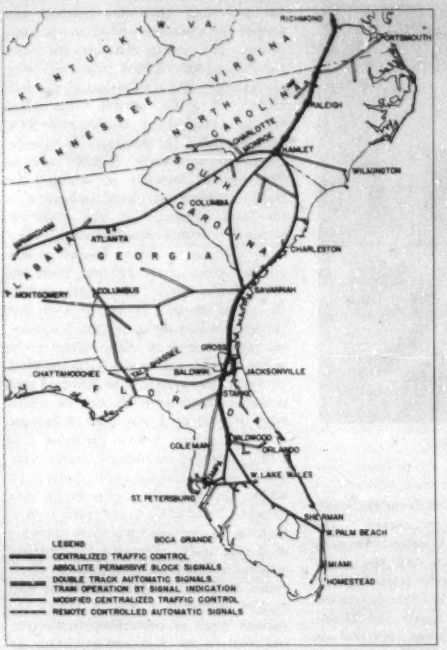


Fig. 6. Various types of centralized traffic control used on Seaboard as of Oct. 1958

machine located in Tampa. Incidentally, at the present time the control line extending between Jacksonville and Miami is the longest centralized traffic-control line in the world to extend in one direction from the control point. Also the system between Richmond and Miami is the longest continuous system in the world on which trains are operated by signal indication.

After the war, in 1946, when materials were again available the Seaboard undertook the installation of traffic control on many of its main lines where signals were not in service. The territories to be controlled were long and a reliable code line was needed. Methods were developed to economically install an underground cable consisting of two no. 10 American Wire Gauge solid copper wires, each circumscribed by 3/64-inch polyethylene insulation. The two polyethylene insulated conductors were twisted with a 4/64-inch Flamenol jacket covering the twisted pair. Details of this installation were given in another paper.[2] Part of this cable has been in service for 11 years, and its performance has been satisfactory with the exception of that cable installed in certain portions of Florida where the ground was infested with pocket gophers. In this territory the pocket gophers, which normally feed on roots, attacked and ate the cable damaging the insulation and eventually causing grounds. The resultant electrolysis eroded the conductors. Experience has proved that the same cable with a 10-mil bronze tape wrapped spirally around the cable affords an armor which protects the cable from pocket gophers. Evidence has been found that pocket gophers gnawed the armored cable but were unable to penetrate the armor. Of all the armored cable installed only one failure has occurred and that as a result of a nail having penetrated the armor and insulation, evidently while on the reel before being installed. It is interesting to note that no lightning failures on the armored cable have occurred but on the nonarmored cable an occasional failure does occur due to lightning, even though a test sample of the cable after 6 years of service was examined and found to have a breakdown voltage in excess of 40,000 volts to ground.

It should be pointed out that with systems of centralized traffic control the wayside signal apparatus used is of high quality. The electric circuits are designed on a "fail safe" basis and, even though the communication facilities controlling such wayside signs may fail to operate properly, the wayside signals are arranged to check that all power-operated switches are mechanically locked, all hand-operated switches are in the proper position, rails are intact, the track ahead is clear for either following or opposing trains, and that the signals governing opposing movements have been at stop for the prescribed predetermined time. In the traffic-control systems installed recently on the Seaboard, in the event of communication failure the train automatically will clear signal in advance if the route ahead is clear. This feature is very desirable and eliminates a large number of delays if a communication failure should occur and it meets the company's first and last requisite which is safety.

References

1. Traffic Control for Railroads, O. W. Baughman. *AIEE Transactions*, vol. 78, pt. II (*Applications and Industry*), July 1959, pp. 173–76.

2. Carrier Operation in Buried Cable, J. R. DePriest. *Proceedings*, Association of American Railroads, New York, N. Y., Communications Sections, Sept. 1948, pp. 374–88.

Traffic Control for Railroads

G. W. BAUGHMAN
FELLOW AIEE

THE HISTORY of the United States shows in some degree a direct relationship between the country's progress and the building of the railroads. One of the railroads' most outstanding achievements was the phenomenal manner in which they handled more than 90% of all personnel and materials that were moved for military purposes during World War II. One objective of this paper is to show how the centrallized traffic-control system helped the railroads achieve this outstanding performance.

Although not generally realized, the railroads have been among the leaders in adopting and developing technological advances. This paper will mention briefly some of the advances which are electrical or electronic in nature. The development of the traffic-control system, which has meant a great deal to the railroads, will then be described.

The railroads adopted the telegraph soon after its invention and developed methods of train operation based upon telegraphic orders. In a similar manner the telephone was adopted. At present the railroads make extensive use of the modern techniques in communications.

The first use of electronic principles on a commercial scale, outside the fields of the telephone and the telegraph, is credited to the automatic cab signal. Development work on this system started in 1916, and the first commercial installation was made in 1923. It is this system which causes a signal to be relayed to the cab indicating the safety of traffic conditions in advance. An important application of this system will be referred to later.

The transition from steam to the diesel-electric locomotive has taken place at a phenomenal rate on the railroads in this country. The change is approaching completion with approximately 27,000 diesel-electric locomotives in contrast to the 3,000 steam locomotives now in service.

The railroads are installing large classification yards for the handling of freight cars. These installations include methods for switching each car to its proper destination track in response to previously stored information. This storage may be on teletype tape or other suitable memory system. Radar devices are used for speed measurement in servo loops for the control of the speeds of the cars. These installations are among the nearest approaches to full automation in service today.

Paper 59-252, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 4, 1958.

G. W. BAUGHMAN is with the Union Switch & Signal, Division of Westinghouse Air Brake Company, Swissvale, Pa.

Description and Discussion

The preceding indicated some of the electrical developments that have been adopted and developed by the railroads. The paper shall now discuss centralized traffic control (CTC). This is actually a system of train operation where the train crew receives no written orders from the dispatcher. The train is operated in accordance with the signal indications. Train orders may be considered as issued by the dispatcher because he is given remote control of the signals and switches in his territory. A basic concept of this system is that it is divded into two main parts. The first is a code communication system so the dispatcher may convey his desires to the signals and switches. Also this communication system transmits information to the dispatcher giving the positions of signals and switches as well as the location of all trains. The second part cousists of a complete wayside signal system with the main signals normally in the "stop" position. The over-all safety resides to a major extent in this part of the system. There are two conditions that must be satisfied before a signal can clear to authorize the passage of a train. The dispatcher must send a code to the particular signal, in effect requesting it to clear. This request is received and decoded automatically and thus the first condition is satisfied. The second condition involves the automatic signal equipment which functions continuously in response to traffic conditions. Before the particular signal can clear, the automatic signaling equipment must check that it is safe for the signal to clear. In a similar manner the dispatcher has control of the track switches in his territory. But before any track switch changes its position, in response to a code from the dispatcher, the automatic signaling equipment must check that it is safe for the switch to move.

Fig. 1 shows a machine at Savannah, Ga., for the control of 250 miles of single-track railroad. A diagram on the machine shows the track layout over which the dispatcher has control. Small lights on the diagram show the positions of all trains and these positions are recorded on a chart usually called an automatic train graph. Also the positions of all switches and signals are indicated by lights on the panel.

The dispatcher sends codes to the various locations in his control of the traffic. This is accomplished merely by the operation of small levers and push buttons. Control codes are generated automatically in response to the positions of the levers and push buttons. Indica-

Fig. 1. A 25-foot control machine in service at Savannah, Ga.

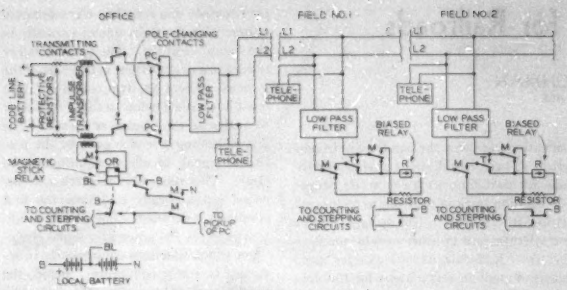


Fig. 2. A multiple CTC line circuit with low-pass filter to permit a physical voice circuit on the same pair of line wires

tion codes from the way stations to the control office are generated in response to the positions of the functions in the field.

Two line wires are used for the transmission of these codes. The use of a single wire with ground return received consideration but was discarded promptly because of the requirement of co-ordination with communications lines. Another reason against the single wire with ground return is the interference that would be experienced during magnetic storms.

The usual installation on a single-track railroad has the field functions arranged in groups of about 6 to 10 functions per station. The stations are spaced along the line at intervals in the order of 5 miles. In usual communications language this might be called a party line. In this application there are frequently 15 to 25 parties on a line.

Fig. 2 shows a simplified line circuit. In the event of simultaneous starts at office and field, the office is given preference by making the first pulse of each control code long as compared with the first pulse of each indication code. The method for establishing priority among the field stations is dependent upon the specific assignment of station codes and the geographical sequence of the stations with respect to the office.

The codes consist of pulses of long and short intervals. The first portion of a code is assigned to station selection and the last portion to the functions of that station.

The main novelty of Fig. 2 with respect to previous line circuits is that the way stations are connected to the line in multiple. The two line wires are continuous throughout the installation whether the line is idle or handling codes. Way stations transmit indication codes by alternately shunting and removing the shunt from the line. This action produces changes in current from the line battery at the office. These changes in current through the indication transformer induce voltages in its relay winding for operation of the office line relay.

Because of the over-all reliability required of this system special attention is given to the strength and ruggedness of the line wires. This being true it is logical to consider the use of these same line wires for other functions. This principle develops in a cumulative manner. The stronger and more rugged the line wires, the greater is the desirability of placing additional functions on that same pair of wires. Also the greater the number of functions being handled over the one pair of wires, the greater is the attention that is warranted in providing reliability and good maintenance.

A major step in permitting additional functions to be handled on the same pair of line wires is the installation of low-pass filters between each way station equipment and the line. The low-pass filter for this service has low resistance to the flow of direct current and high impedance to frequencies in the voice and ultrasonic ranges. The filter is designed to minimize the electrical noise which is fed to the line because of the coding contacts. To maintain ruggedness and reliability of the line circuit, special attention is given to this filter so that it will withstand the severe surges that develop on lines of this type during storms. This special attention to the filter does much to protect other apparatus at the station from damage by lightning surges. With these low-pass filters in place, as shown in Fig. 2, the pair of wires for the CTC line serves as an excellent voice communications circuit. It is the usual practice to have a telephone at each way station so that maintenance men and train crews may communicate directly with the dispatcher.

The system as described up to this point was available in 1940, largely because of the intelligent use of funds and personnel for intensive development work during the depression years of the Thirties. During the early days of World War II, the traffic congestion on the single-track railroads was so great that urgent studies were made to determine the best way to relieve the congestion promptly and consistent with the economic use of strategic materials. The traffic control system was determined as the most feasible solution. Emergency measures and much overtime work by the manufacturers and the installation crews provided relief at the most critical locations.

The system operated satisfactorily in practice, but soon after there was a demand for further expansion beyond the approximately 75-mile limit of the system as it existed. The experience indicated that one dispatcher could handle much more territory and that additional advantages would be realized with larger concentration of controls.

The solution to permit large installa-

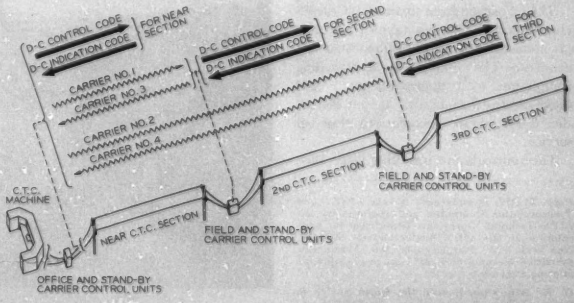


Fig. 3. How carrier frequencies permit the expansion of the CTC control systems

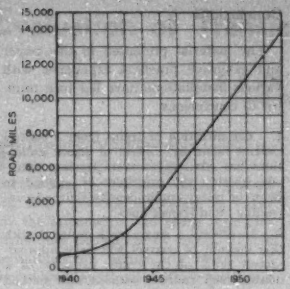


Fig. 4. Total road miles of traffic control versus years

tions, hundreds of miles in length, to be operated from one control office gave full consideration to the desirability of retaining as standard the system and equipment that had been giving excellent service. After a great deal of study of various approaches to a solution, the one that proved to be most acceptable considered the system as it then existed as a unit of a larger system.

Fig. 3 shows three such units placed end to end and controlled from one office. The near unit operates in the conventional manner, previously described. The second unit is arranged in exactly the same manner as if there were a control point located at the far end of the first unit. One carrier frequency transmitted in each direction over the first unit permits the second unit to be controlled over the line wires of the first. In a similar manner a pair of carrier frequencies over the line wires of the first and second units permits the handling of a third unit. The filters which were provided to permit a telephone communication system to be operated on the one pair of wires also serve to permit the transmission of the carrier frequencies. This use of carrier frequencies removed the previous limitations as to size of the installations. Railroads may now handle any desired size of installation from one control point. An outstanding example is the Savannah office on the Seaboard Air Line railroad from which over 600 miles of track are controlled.

To retain the telephone circuit on the same pair of line wires it was necessary to select the frequencies of the carriers so that they would not be audible to the extent of causing interference or annoyance to telephone conversations. Tests and experience indicated 11 kc to be the minimum frequency that was practical. It is interesting to note that the first carrier units for this type of service were built on a temporary basis to serve until more substantial units could be made. These units were installed on the Southern Pacific in 1942 and are reported to be continuing in satisfactory service at the present time. To minimize the delay to trains that might be caused by the failure of electronic components the carrier units were installed in duplicate.

Starting with the first carrier units there was an attempt to simplify the manufacturing and inventory problems by making as large a portion of the unit as possible common to all frequencies. The frequency determining units were made separately as plug-in devices.

Special attention should be given to the way that the use of carrier frequencies expanded the size and scope of CTC. All of the main items retained the same use as in the previous system. The addition of carrier frequency and coordinating equipment broke through the previous barriers of the size of installation and the distance of transmission. This indicates why it was possible to meet the sudden demands for larger installations. An idea of the sudden change in demand is indicated by noting the change in slope of the curve in Fig. 4 during the years 1942 to 1943.

An important application of the equipment and system just described, outside the railroad field, was in the extensive remote control and indication system for some of the vital experiments in the Manhattan Project.

An interesting demonstration was performed jointly by the Pennsylvania Railroad, the Radio Corporation of America, Western Union, and Union Switch & Signal in 1946 to show that the control office could be separated from the active territory by great distances and various forms of communications. The transmission path between the office and the active territory in the demonstration included carrier frequencies on railway line wires, Western Union telegraph lines, and several sections of microwave.

The 90-mile CTC installation which was made on the Orinoco Project in Venezuela, with the co-operation of the Radio Corporation of America, in 1953, is novel with respect to all other installations because it uses microwave principles instead of line wires.

A recent and significant development has been made in the control machine to facilitate the work of the dispatcher. Fig. 5 shows this new concept of machine which has the track diagram compressed in the ratio of approximately 4 to 1 with respect to previous machines. This was made possible by removing the push buttons and levers from direct physical relationship on the panel to the functions being controlled. A relatively small push-button panel is located on the desk portion of the machine. By operation of the push buttons on this one panel the dispatcher may issue controls to any desired way station. A logical arrangement in the numbering of way stations and functions permits the operator to issue his controls without the need of reference to a chart or the memorizing of code letters or numbers.

The material which has been presented thus far applies mainly to systems for the handling of traffic on single-track lines of the Western and Southern parts of our country. These same general principles have been applied in the reduction from

Fig. 5. New concept in traffic control machines

two tracks to a single track on some railroads. In some other instances there are reductions from three and four tracks to two tracks.

When the broad principles of CTC are applied to a railroad with large traffic density, such as the rapid transit lines in New York City, it is necessary to make important changes in the code control and indication system. Because of the relatively short distances involved there is not the same urgency to place many facilities on one pair of wires. There is the requirement for the transmission of more information and in less time. This is accomplished by using systems with greater speed and fewer stations. The latest systems permit controls to the field simultaneously with indications to the office.

This brief reference to the systems on rapid transit lines is made to emphasize the range of application of the basic principle of directing trains by signal indications and to point out that systems have been developed to implement this principle whether it be on a small stretch of track in a remote area, hundres of miles of medium traffic density, or on the dense traffic of rapid transit lines.

In the advancement of these developments there were always definite goals set for the purpose of steering the developments along logical lines. One important goal that is being taken into account at present, and in part guiding development programs, is the crewless train. It is obvious that CTC as discussed in this paper constitutes one of the important components in the operation of crewless trains. The system transmits the desires of the dispatcher to the signal system and transmits information regarding the signals, switches, and trains back to the dispatcher. By the techniques used in cab signaling, the control information from the dispatcher may be applied to the rails and transferred inductively to the cab of the locomotive. Relays on the locomotive controlled in this manner by the dispatcher may have contacts in the locomotive control circuits. Thus the dispatcher would have remote control of the locomotive and safety of operation would be provided by the way side and cab-signaling equipment.

Conclusions

The material presented in this paper and extensive practical experience with various installations have led to a number of conclusions.

1. The direction of trains from a central control point by control of signal indications has advantages over previous methods whether the principles be applied to light- or heavy-traffic density railroads, including rapid transit lines.

2. In general a single-track railroad with this method of control will have approximately 80% of the traffic capacity of a double-track railroad without this method of control.

3. After CTC has been installed on a single-track railroad, the improvement in the movement of a freight train is indicated by a saving of approximately one minute for each mile of length of the installation.

4. The principles of CTC are being applied in some foreign countries to expand the traffic capacity of existing railroads or are being included in the construction programs of new railroads.

5. From a practical standpoint the time may be quite near when crewless trains are operated on some rapid transit lines.

Evaluation of Transient System Response

F. P. DE MELLO
MEMBER AIEE

FREQUENCY-RESPONSE techniques are among the most powerful tools in the hands of the feedback control systems engineer. Perhaps of greatest use today are the attenuation and phase-angle plot techniques developed by Bode. These techniques are particularly effective in analysis and synthesis of stable systems in establishing the relative stability of a system.[1] Although it is possible to estimate the transient response of a system whose open-loop frequency response is known, the accurate prediction of this transient response by analytical means involves obtaining the roots of the characteristic equation for the closed-loop system. Knowing the system's closed-loop factored transfer function, the transient response of the system may then be obtained analytically by means of Laplace transform methods. Thus, if

$$S(s)=\frac{C(s)}{R(s)}$$

is the system's closed-loop transfer function, then the Laplace transform of the output $C(s)$ for a reference step of magnitude $A(R(s)=A/s)$ would be $C(s)=(A/s)S(s)$.

The only difficulty with the application of Laplace transform methods lies in the fact that, in order to obtain the inverse time solution, the roots of the denominator of the closed-loop transfer function

$$\frac{G(s)}{1+G(s)H(s)}$$

must be determined. Although the function $G(s)H(s)$ is generally known in terms of its poles and zeros, the expression $1+G(s)H(s)$ usually turns out to be a high-order polynomial in s whose roots are not easily extracted.

This paper illustrates a straightforward method of obtaining the closed-loop system transfer function in terms of an $\mathcal{L}$ transform expression whose inverse may be readily obtained. This method is based in part on some ideas which were presented in a discussion by A. Leonhard.[2] It is a logical extension of attenuation and phase-angle plot techniques using damped frequencies. It is considerably easier than Root Locus methods[3] and has much appeal to the engineer who is trained in Bode's attenuation and phase-angle techniques.

The application of the method developed in this paper is illustrated with examples.

In the analysis and synthesis of systems it often is necessary to represent into factored transfer function form inner loops, or sections which may be known as a summation of transfer functions. The method of obtaining roots outlined in this paper permits expressing any part of a system, made up of components in feedback or in summation, into factored transfer function form.

The Problem

The general problem can be stated as one of expressing the transfer function of a closed loop in terms of a factored numerator and denominator.

Thus, if $G(s)H(s)$ is the open-loop transfer known in terms of its poles and zeros

$G(s)$ the forward function

$$=\frac{(s+a_1)(s+a_2)(s+a_3)\dots}{(s+b_1)(s+b_2)(s+b_3)\dots} \quad (1)$$

$H(s)$ the feedback function

$$=\frac{(s+c_1)(s+c_2)(s+c_3)\dots}{(s+d_1)(s+d_2)(s+d_3)\dots} \quad (2)$$

The closed-loop transfer function is

$$\frac{G(s)}{1+G(s)H(s)}=\frac{(s+a_1)(s+a_2)\dots(s+d_1)(s+d_2)\dots}{(s+a_1)(s+a_2)\dots(s+c_1)(s+c_2)\dots\ (s+b_1)(s+b_2)\dots(s+d_1)(s+d_2)\dots} \quad (3)$$

The numerator of the expression contains the same factors as appear in the numerator of the forward function $G(s)$ multiplied by those in the denominator of the feedback function $H(s)$.

The factors in the denominator are the roots of $(1+GH)$ and can be expressed as

$$(s+r_1)(s+r_2)(s+r_3)\dots \text{ where } r_1, r_2, r_3\dots$$

are real or complex roots of $1+GH=0$, i.e., values of s for which $G(s)H(s)=-1$.

The determination of these roots is all that is required to obtain the transient response of a system in terms of a Laplace transform expression whose inverse or time solution can be easily extracted.

As a corollary, if the summation of two functions $G_1(s)+G_2(s)$ has to be represented into an over-all factored function $G_3(s)$, and if

$$G_1(s)=\frac{(s+a_1)(s+a_2)(s+a_3)}{(s+f_1)(s+f_2)(s+f_3)} \quad (4)$$

and

$$G_2(s)=\frac{(s+g_1)(s+g_2)(s+g_3)}{(s+h_1)(s+h_2)(s+h_3)} \quad (5)$$

Then

$$G_3(s)=\frac{K[(s+i_1)(s+i_2)(s+i_3)\dots]}{(s+f_1)(s+f_2)(s+f_3)(s+h_1)\times(s+h_2)(s+h_3)} \quad (6)$$

where K is a constant which may be readily determined.

The roots of the following expression are i_1, i_2, i_3, etc.

$$1+\frac{G_1(s)}{G_2(s)}=0$$

i.e., values of s which make

$$\frac{G_1(s)}{G_2(s)}=-1$$

Determination of Roots

A polynomial in s will have real roots and/or pairs of conjugate complex roots. For a polynomial in s describing the characteristic equation of a stable system, the real roots or real parts of the complex roots will always be negative.

Thus the roots of a stable system can only be of the two forms

$$s_0=-\alpha_0 \quad (7)$$

and

$$s_{1,2}=-\alpha\pm j\omega \quad (8)$$

A polynomial having such roots could be expressed, in factored form, as

$$(s+\alpha_0)[(s+\alpha)^2+\omega^2] \quad (9)$$

The techniques of attenuation and phase-angle plots versus frequency, so effectively and widely used for system synthesis and analysis, can be adapted for use with complex frequencies. The procedure of locating the roots can then be stated as the search for the real or complex frequencies for which the magnitude of the function $G(s)H(s)$ is equal to 1 and its phase angle is 180 deg (degrees) or odd multiples of 180 deg.

Real Roots of Closed Loop

Examine the function

$$F(s)=\frac{1}{(1+sT)} \quad (10)$$

Paper 59-255, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted August 29, 1958; made available for printing December 3, 1958.

F. P. DE MELLO is with the General Electric Company, Schenectady, N. Y.

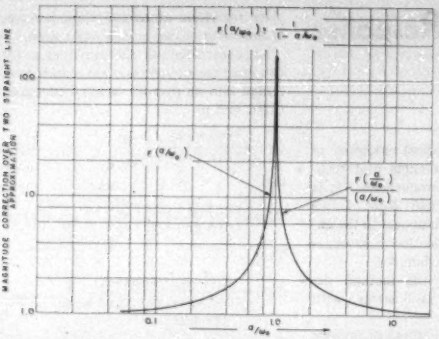


Fig. 1. Magnitude corrections over two straight-line Bode approximation for function F(S), with $s=-\alpha$ plotted as function of α/ω_0

$$F(s) = \frac{1}{1+\frac{s}{\omega_0}}$$

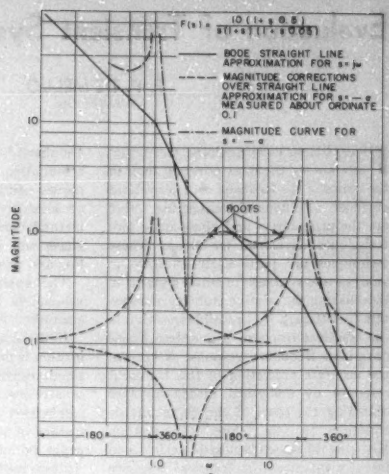


Fig. 2 (right). Bode plot of open-loop function given in curve

for $s=-\alpha$,

$$F(-\alpha)=\frac{1}{1-\alpha T} \tag{11}$$

$$=\frac{1}{\left[1-\frac{\alpha}{\omega_0}\right]} \tag{12}$$

where $\omega_0=1/T$ is known as the break frequency in Bode plot language.

The magnitude plot of $F(-\alpha)$ has asymptotic values of

$$|F(-\alpha)| = 1 \text{ for } \frac{\alpha}{\omega_0} << 1 \text{ and } |F(-\alpha)| = \frac{1}{\frac{\alpha}{\omega_0}} \text{ for } \frac{\alpha}{\omega_0} >> 1$$

As α/ω_0 approaches 1.0, the magnitude of the function increases rapidly and becomes infinite at $\alpha/\omega_0=1.0$. It decays rapidly from infinity as α/ω_0 becomes greater than 1.0.

The phase angle of the function $F(-\alpha)$ is 0 deg from $0<\alpha/\omega_0<1$ and is 180 deg for $\alpha/\omega_0>1.0$.

One easy way of finding the real roots of a closed-loop system from a Bode plot of its open-loop function is to prepare a template for the function $F(-\alpha/\omega_0)$ giving the magnitude correction over the two straight-line asymptotic approximation for real frequencies, as shown in Fig. 1.

Fig. 2 illustrates an example of solution for the real roots of a system whose forward function is $1/s(1+s)$ and feedback function is $10(1+s0.5)/(1+s0.05)$. The over-all open-loop function is $10(1+s0.5)/s(1+s)(1+s0.05)$.

The break frequencies for the denominator are $\omega_1=1.0$, $\omega_2=20$.

The numerator has a lead factor with a break frequency of $\omega_1'=2.0$.

On the log-magnitude versus log-frequency plot of the function's straight-line approximation, we have also drawn in the magnitude corrections for real frequencies. This was done with the aid of the real frequency template which was placed with its $\alpha/\omega_0=1.0$ marker opposite each break frequency. Note that the corrections for a lead factor in the numerator are a mirror image of the corrections for a lag factor in the denominator. For convenience, these component corrections were drawn about the ordinate 0.1. A negative correction is measured as the distance of the component curve below the ordinate 0.1. A positive correction would be the distance of component curve above the ordinate 0.1. The total magnitude correction to be applied to the straight-line approximation would be obtained by summing all component corrections which can be conveniently done with a pair of dividers.

It is also a simple matter to establish the regions where the function has a phase angle of 0 deg or even multiples of 180 deg and those regions where the function has a phase angle of 180 deg and/or odd multiples of 180 deg. Wherever the Bode straight-line approximation has slope of one or odd multiples, the phase angle for real frequencies is 180 deg or odd multiples of 180 deg. Crossovers of the magnitude $GH(-\alpha)$ with the ordinate 1 represent real roots of $GH+1=0$.

From Fig. 2, the roots of $1+GH$ are

$s_1=-3.10$
$s_2=-5.0$
$s_3=-12.9$

The transfer function of the closed loop can be written as

$$F(s)=\frac{\left(1+\frac{s}{20}\right)}{10\left(1+\frac{s}{3.10}\right)\left(1+\frac{s}{5.0}\right)\left(1+\frac{s}{12.9}\right)}$$

or in the form more readily adaptable for the inverse transformation

$$F(s)=\frac{3.10(5.0)(12.9)(s+20)}{10[20(s+3.1)(s+5.0)(s+12.9)]}$$

The response for a unit step input would be $C(s)=F(s)/s$

$$=\frac{3.10(5.0)(12.9)(s+20)}{10[20s(s+3.1)(s+5.0)(s+12.9)]}$$

The time response would then be

$$C(t)=K+K_1e^{-3.10t}+K_2e^{-5.0t}+K_3e^{-12.9t}$$

where

$$K=\frac{3.10(5.0)(12.9)(20)}{10(20)(3.10)(5.0)(12.9)}=0.10$$

$$K_1=\frac{3.10(5.0)(12.9)(20-3.10)}{10(20)(-3.10)(5.0-3.10)(12.9-3.10)}$$
$$=-0.293$$

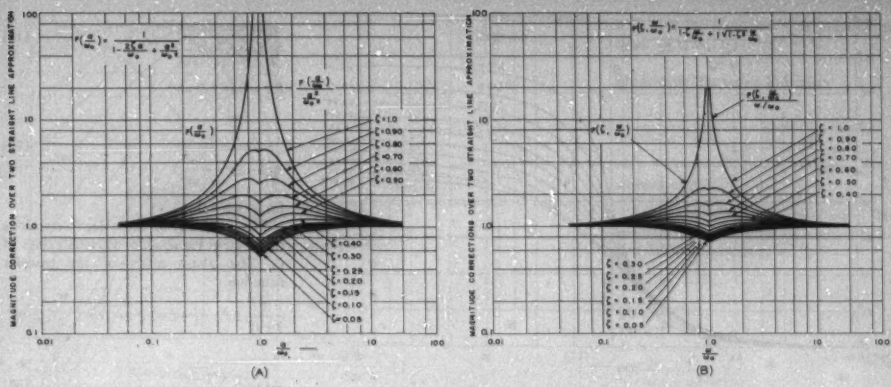

MAGNITUDE CORRECTION OVER TWO STRAIGHT LINE APPROXIMATION
MAGNITUDE CORRECTIONS OVER TWO STRAIGHT LINE APPROXIMATION
(A)
(B)

Fig. 3. Magnitude corrections over two straight-line Bode approximation for function F(S)

A—$s=-\alpha$ plotted as function of α/ω_0

B—Plotted as function of ω/ω_0 for various values of ζ

$$F(s)=\frac{1}{1+\frac{2\zeta s}{\omega_0}+\frac{s^2}{\omega_0^2}}$$

$$F(s)=\frac{1}{1+\frac{s}{\omega_0}} \text{ with } s=(-\zeta+j\sqrt{1-\zeta^2})\omega$$

$$K_2=\frac{3.10(5.0)(12.9)(20-5.0)}{20(-5.0)(3.10-5.0)(12.9-5.0)}$$

$$=0.200$$

$$K_3=\frac{3.10(5.0)(12.9)(20-12.9)}{20(-12.9)(3.10-12.9)(5.0-12.9)}$$

$$=0.0071$$

$$C(t)=0.10-0.293e^{-3.10t}+0.200e^{-5.0t}-0.007e^{-12.9t}$$

A factor in the open-loop function containing complex roots is usually expressed in the form

$$F(s)=\frac{1}{1+\frac{2\zeta}{\omega_0}s+\frac{s^2}{\omega_0^2}} \tag{13}$$

where the roots are

$$s_{1,2}=(-\zeta\pm j\sqrt{1-\zeta^2})\omega_0$$

$$0<\zeta<1$$

For $s=-\alpha$

$$F(-\alpha)=\frac{1}{\left[1+\frac{\alpha^2}{\omega_0^2}-\frac{2\zeta}{\omega_0}\alpha\right]} \tag{14}$$

Since $0<\zeta<1.0$, this function is always positive (phase angle=0 deg) and its magnitude is asymptotic to 1 at $\alpha/\omega_0<<1$ and asymptotic to

$$\frac{1}{\frac{\alpha^2}{\omega_0^2}} \text{ at } \frac{\alpha}{\omega_0}>>1$$

with a resonant peak at

$$\frac{\alpha}{\omega_0}=\zeta \text{ equal to } \frac{1}{1-\zeta^2}$$

Curves of the magnitude of log $F(-\alpha)$ minus log of the two straight-line approximation versus α/ω_0 are shown on Fig. 3(A).

Since such complex root factors do not contribute any phase shift, their effect on real roots is merely by changing the magnitude curve and thereby possibly changing the point of crossover.

Complex Roots of Closed Loop

Pairs of conjugate complex roots

$$(-\zeta\pm j\sqrt{1-\zeta^2})\omega$$

are the other possibility. It suffices to test for one of the conjugates, say

$$(-\zeta+j\sqrt{1-\zeta^2})\omega$$

and find the complex frequency in terms of the parameters ζ and ω for which the magnitude of the function $GH(\zeta,\omega)=1.0$ and its phase angle is equal to 180 deg or odd multiples of 180 deg.

The magnitude and phase-angle behavior of the common transfer functions when excited by damped frequencies will be discussed in the following. A transfer function

$$F(s)=\frac{1}{1+sT} \tag{15}$$

for $s=[-\zeta+j\sqrt{1-\zeta^2}]\,\omega$ becomes

$$\frac{1}{1-\zeta\frac{\omega}{\omega_0}+j\sqrt{1-\zeta^2}\frac{\omega}{\omega_0}} \tag{16}$$

where $\omega_0=1/T$ is ordinarily known as break frequency in Bode asymptotic techniques.

Since the magnitude of the complex frequency

$$|(-\zeta+j\sqrt{1-\zeta^2})\omega|=\omega$$

it is evident that a plot of the log of the magnitude of the function

$$F[(-\zeta+j\sqrt{1-\zeta^2})\omega]$$

versus the log of the magnitude of complex frequency ω will be asymptotic to log 1 at $\omega/\omega_0<<1.0$ and asymptotic to log

$$\frac{1}{\omega/\omega_0} \text{ at } \frac{\omega}{\omega_0}>>1.0$$

and therefore the complex break frequency which is the intersection of the two asymptotes will again occur at ω_0. The actual magnitude is

$$\frac{1}{\sqrt{1-2\zeta\frac{\omega}{\omega_0}+\frac{\omega^2}{\omega_0^2}}} \tag{17}$$

It is interesting that this expression is equal to the square root of the magnitude of a quadratic transfer function when excited by a real frequency α, equation 14. The ratio between the actual magnitude and the two straight-line approximation is

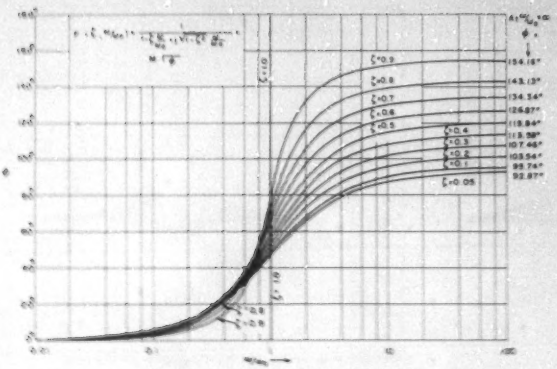


Fig. 4. Angle of function plotted as function of ω/ω_n for various values of ζ

$$\frac{1}{\sqrt{1-2\zeta\frac{\omega}{\omega_n}+\frac{\omega^2}{\omega_n^2}}} \tag{18}$$

for $\omega/\omega_n < 1.0$ and

$$\frac{\frac{\omega_n}{\omega}}{\sqrt{1-2\zeta\frac{\omega}{\omega_n}+\frac{\omega^2}{\omega_n^2}}} \tag{19}$$

for $\omega/\omega_n > 1.0$.

Fig. 3(B) shows these log corrections as function log ω/ω_n for various values of ζ. The phase angle of the transfer function

$$\frac{1}{1+s\tau} \text{ for } s=\left[-\zeta+j\sqrt{1-\zeta^2}\right]\omega_n$$

is

$$-\tan^{-1}\frac{\sqrt{1-\zeta^2}}{\frac{\omega_n}{\omega}-\zeta} \tag{20}$$

From this expression the following relations are apparent.

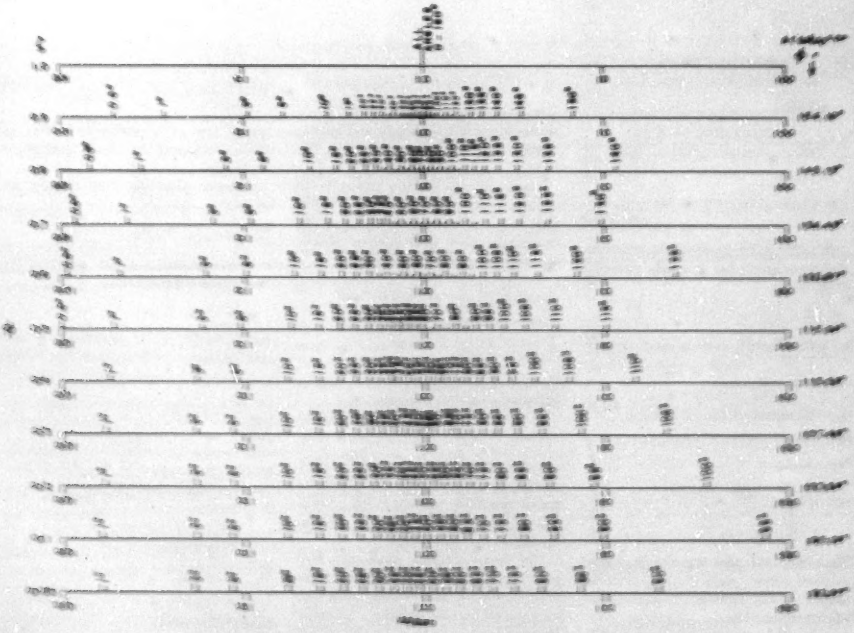


Fig. 5. Phase-angle values giving angle of (18) as function of ω/ω_n for various values of ζ

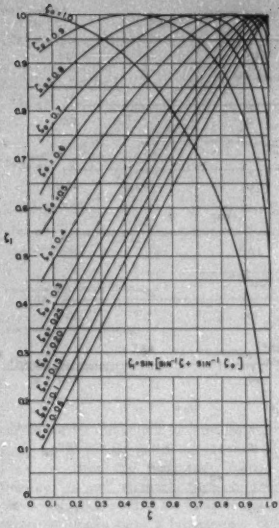


Fig. 6. Plots of ζ_1 as function of ζ for various values of ζ_0

ω/ω_0	ϕ, Degrees
$\frac{1}{\zeta+\sqrt{1-\zeta^2}}$	-45
$\frac{1}{\zeta}$	-90
$\frac{1}{\zeta-\sqrt{1-\zeta^2}}$	-135

$$\tan\phi = \frac{\sqrt{1-\zeta^2}}{\frac{\omega_0}{\omega}-\zeta} \qquad (21)$$

or

$$\frac{\omega}{\omega_0} = \frac{1}{\sqrt{1-\zeta^2}\tan\phi+\zeta} \qquad (22)$$

for $\omega/\omega_0 = 1.0$

$$\phi = \tan^{-1}\frac{\sqrt{1+\zeta}}{\sqrt{1-\zeta}} \qquad (23)$$

Fig. 4 shows values of angle as function of ω/ω_0 for various values of ζ. Phase-angle rules for various values of ζ can be conveniently made up as shown on Fig. 5 to give the phase angle as function of ω/ω_0.

Quadratic Transfer Function

$$\frac{1}{1+2\zeta_0 T_0 s+T_0^2 s^2} \text{ or } \frac{1}{1+\frac{2\zeta_0}{\omega_0}s+\frac{s^2}{\omega_0^2}}$$

where $0<\zeta_0<1$

Let us now observe the response of the resonant quadratic function when excited by a damped frequency,

$(-\zeta+\sqrt{1-\zeta^2})\omega$

The function

$$\frac{1}{1+\frac{2\zeta_0}{\omega_0}s+\frac{s^2}{\omega_0^2}} \qquad (24)$$

can be written as

$$F(s) = \frac{1}{\left[1+\frac{s}{(\zeta_0+j\sqrt{1-\zeta_0^2})\omega_0}\right]\times\left[1+\frac{s}{(\zeta_0-j\sqrt{1-\zeta_0^2})\omega_0}\right]} \qquad (25)$$

With $s=[-\zeta+\sqrt{1-\zeta^2}]\omega$

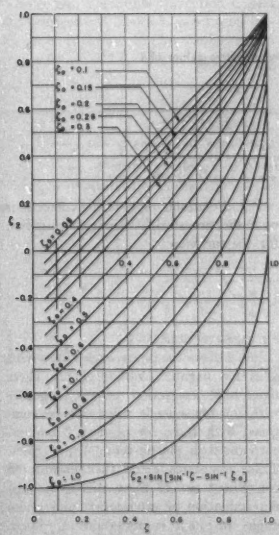


Fig. 7. Plots of ζ_2 as function of ζ for various values of ζ_0

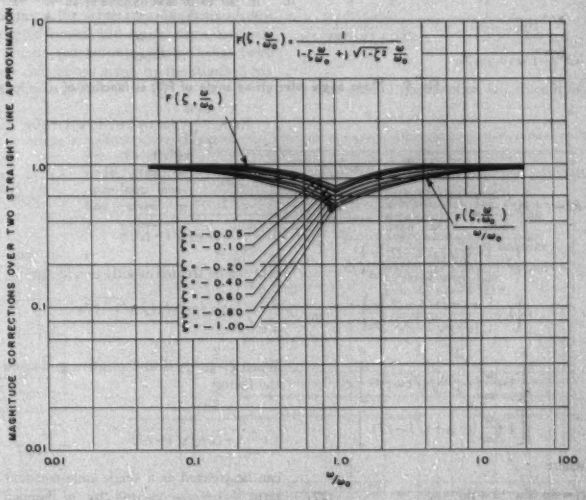


Fig. 8. Magnitude corrections over two straight-line approximation for the function F(S), and negative values of ζ

$$F(s) = \frac{1}{1+\frac{s}{\omega_0}} \text{ with } s=(-\zeta+j\sqrt{1-\zeta^2})\omega$$

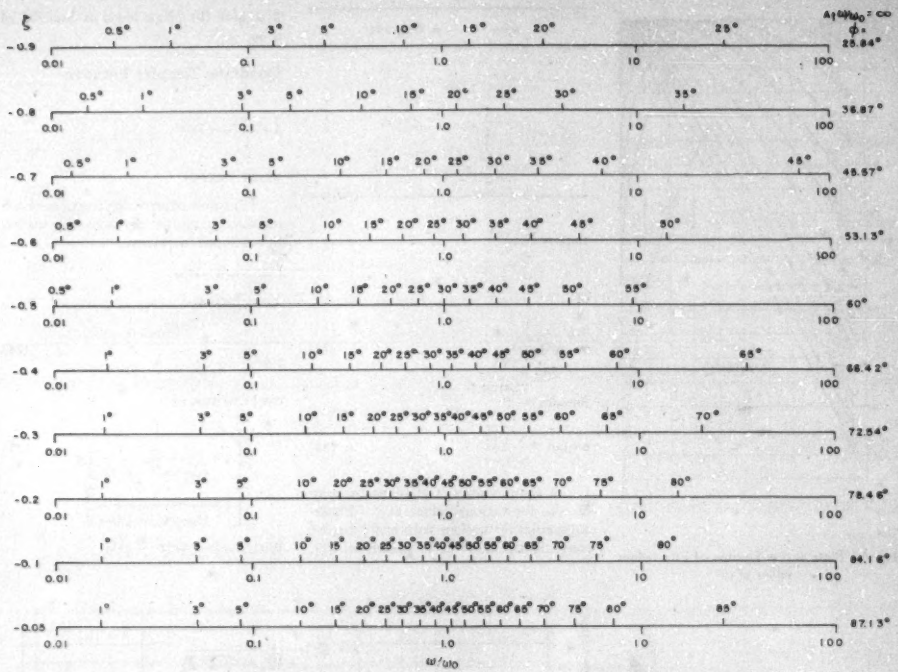


Fig. 9. Phase angle rules giving angle of F(S) as function of ω/ω_0 for various negative values of ζ

$$F(s)=\frac{1}{1+\frac{s}{\omega_0}} \quad \text{for } s=(-\zeta+j\sqrt{1-\zeta^2})\omega$$

$$F(-\zeta+j\sqrt{1-\zeta^2})\omega = \frac{1}{\left\{1+\left[\frac{-\zeta+j\sqrt{1-\zeta^2}}{\zeta_0+j\sqrt{1-\zeta_0^2}}\right]\frac{\omega}{\omega_0}\right\}\times \left\{1+\left[\frac{-\zeta+j\sqrt{1-\zeta^2}}{\zeta_0-j\sqrt{1-\zeta_0^2}}\right]\frac{\omega}{\omega_0}\right\}} \quad (26)$$

$$= \frac{1}{\left[1+\frac{\omega}{\omega_0}(-\zeta_1+j\sqrt{1-\zeta_1^2})\right]\times \left[1+\frac{\omega}{\omega_0}(-\zeta_2+\sqrt{1-\zeta_2^2})\right]}$$

where

$$\zeta_1=\sin\left[\sin^{-1}\zeta+\sin^{-1}\zeta_0\right] \quad (27)$$

$$\zeta_2=\sin\left[\sin^{-1}\zeta-\sin^{-1}\zeta_0\right] \quad (28)$$

Figs. 6 and 7 are plots of ζ_1 and ζ_2 as function of ζ for various values of ζ_0. ζ_1 is always positive for values of ζ and ζ_0 from 0 to 1.0, and the factor

$$\frac{1}{1+\frac{\omega}{\omega_0}(-\zeta_1+j\sqrt{1-\zeta_1^2})}$$

and can be treated exactly as a factor

$$\frac{1}{1+\frac{s}{\omega_0}} \text{ with } s=-\zeta_1+j\sqrt{1-\zeta_1^2}\omega$$

When $\zeta>\zeta_0$, ζ_2 is also positive and again the factor

$$\frac{1}{1+\frac{\omega}{\omega_0}(-\zeta_2+j\sqrt{1-\zeta_2^2})}$$

can be treated as a single time constant term $1/1+s/\omega_0$ excited by a damped frequency

$$(-\zeta_2+j\sqrt{1-\zeta_2^2})$$

when $\zeta<\zeta_0$, ζ_2 is negative, which is equivalent to having the function $1/1+s/\omega_0$ excited by

$$s=(\zeta_2+j\sqrt{1-\zeta_2^2})\omega$$

Fig. 8 contains plots of log corrections over the two straight-line approximation for the function

$$\frac{1}{1+\frac{\omega}{\omega_0}(-\zeta+j\sqrt{1-\zeta^2})}$$

for negative values of ζ.

Fig. 9 gives the phase-angle rule versus ω/ω_0 for negative values of ζ. These magnitude correction curves, together with the phase-angle rules, can be combined with those for positive ζ and incorporated in plastic templates or rules. These templates would then permit the convenient use of these curves for the determination of attenuation and phase angle of any combination of open-loop transfer functions from which the closed-loop roots can be easily extracted.

The essential magnitude correction

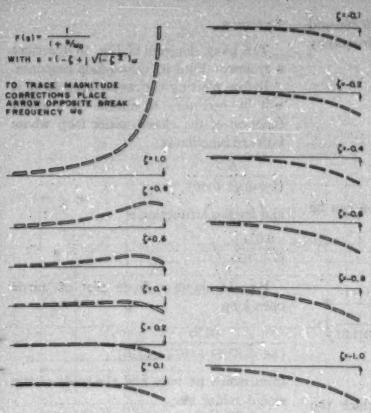


Fig. 10. Model of template for magnitude corrections for different damping ratios ζ

curves can be presented in templates or transparent plastic rulers as shown on Fig. 10. Here the magnitude correction curves are stamped through the plastic template so that they can be conveniently drawn onto a log-log plot by guiding a sharp pencil through the narrow gap along the curve. The phase-angle rules shown on Figs. 5 and 9 can also be presented in transparent plastic rulers for convenient use.

Since the magnitude correction curves are symmetrical, about $\omega/\omega_0 = 1.0$ on log-log plots, it suffices to plot the correction curves for $\omega/\omega_0 < 1.0$. By turning over the template, the same curve would apply for $\omega/\omega_0 > 1.0$.

The magnitude corrections for a complex quadratic factor of damping ratio ζ excited by real frequencies is equal to twice the magnitude correction of a normal time constant factor excited by a damped frequency with same damping ratio ζ. It is therefore unnecessary to include in a template the magnitude corrections for a complex quadratic factor excited by real frequencies since these corrections can be obtained by using twice the corrections given by the damped frequency curves.

On another template could also be included the curves of Figs. 6 and 7 relating ζ_1 and ζ_2 to ζ and ζ_0. These are of use in dealing with quadratic complex factors as explained in the text.

These templates, together with the phase-angle rules of Figs. 5 and 9, would be all that is needed for determination of roots of any system made up of components whose factored transfer functions are known.

EXAMPLE

Take the system whose forward function is given by the expression

$$\frac{0.10}{s(1+0.5s)}$$

and the feedback function by the expression

$$\frac{(1+s20)}{(1+s3.33)}$$

The plot of the open-loop function GH is shown on Fig. 11. The closed-loop response function would evidently be

$$\frac{(1+s3.33)}{(\quad)(\quad)(\quad)}$$

where the factors of the denominator are the roots of $GH+1$, to be determined from the magnitude and phase-angle plot of

$$\frac{0.1(1+20s)}{s(1+0.5s)(1+3.33s)} \quad \text{for } s = -\alpha \text{ and } s = (-\zeta + j\sqrt{1-\zeta^2})\omega$$

The gain of the function is 1 since it is

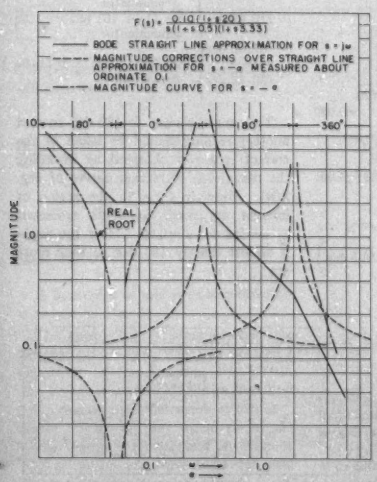

BODE STRAIGHT LINE APPROXIMATION FOR s = jω
MAGNITUDE CORRECTIONS OVER STRAIGHT LINE APPROXIMATION FOR s = −α MEASURED ABOUT ORDINATE 0.01
MAGNITUDE CURVE FOR s = −α
REAL ROOT
MAGNITUDE

Fig. 11 (left). Bode plots of open-loop function given in curve

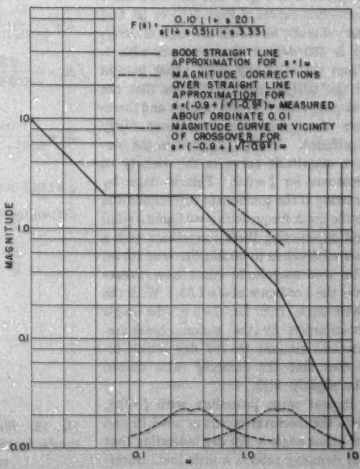

BODE STRAIGHT LINE APPROXIMATION FOR s = jω
MAGNITUDE CORRECTIONS OVER STRAIGHT LINE APPROXIMATION FOR s = (−0.9 + j√(1−0.9²))ω MEASURED ABOUT ORDINATE 0.01
MAGNITUDE CURVE IN VICINITY OF CROSSOVER FOR s = (−0.9 + j√(1−0.9²))ω
MAGNITUDE

Fig. 12 (right). Bode plot of open-loop function given in curve

an integrating system and the numerator contains the denominator of H as a clearing of fractions operation of the function

$$\frac{G}{1+GH}$$

will indicate.

Scanning for Real Roots

Using the real-root template, the magnitude curve of $GH(s)$ can be drawn for $s=-\alpha$ as function of α. Fig. 11 shows this curve as composed from the corrections corresponding to each break which can be added with a pair of dividers to give the total correction. By inspection it is evident that the single slope regions, whether upward or downward, are the regions of 180-deg phase shift. (Odd multiples of such slopes are also regions of 180-deg phase shift.) Crossovers of the magnitude of $GH(-\alpha)$ with the ordinate 1 do give real roots only in these regions. The small portions of the magnitude curves of interest are quickly spotted and these regions can be drawn in with accuracy to obtain the roots.

On Fig. 11 it is seen that in the first section of 180-deg phase shift there is a crossover at 0.035 radian indicating a negative real root $=-0.035$. The next crossovers occur in regions of 0-deg phase shift and therefore there are no further real roots.

Complex Roots

Using the phase-angle rule for damped frequencies of Fig. 5 and the magnitude correction curves of Figs. 3(B) or 10, find the value of ζ for which the angle of $GH(\zeta, \omega)$ is 180 deg at the crossover with 1. From the relatively large phase margin of the $GH(j\omega)$ plot, it follows that the damping ratio ζ should be high and therefore the magnitude corrections fairly significant. In Fig. 12, draw in the correct magnitude curve in the vicinity of crossover for $\zeta=0.9$. This is done by drawing in the component corrections due to the break frequencies $\omega=0.3$ and $\omega=2.0$ and superposing these corrections with a pair of dividers to obtain the true curve. The true magnitude curve now crosses over the 1 ordinate at $\omega=1.55$. With the phase-angle rule for $\zeta=0.9$, the angle contribution by the various break frequencies amount to 187 deg which indicates that the damping ratio chosen was slightly high.

Repeat same procedure with $\zeta=0.8$. Here the crossover occurs at $\omega=1.10$ and the phase angle is 160 deg indicating that this damping ratio is a little low. Interpolation will yield a value of $\zeta=0.866$ and $\omega=1.31$ as the correct complex frequency for which

$$GH[-\zeta+j\sqrt{1-\zeta^2}]_{\omega=-1}$$

i.e., the required complex root.

The roots then are

$$s=-0.035$$
$$s=(0.866\pm j0.50)\ 1.31$$

The closed-loop transfer function can be written as

$$\frac{(G)s}{1+G(s)H(s)} = \frac{(1+s3.33)}{\left(1+\frac{s}{0.035}\right)\left(1+\frac{s}{(0.866+j0.50)1.31}\right)\times \left(1+\frac{s}{(0.866-j0.50)1.31}\right)}$$

Expressed into a form from which the inverse Laplace transform can be readily extracted

$$\frac{G(s)}{1+G(s)H(s)} = \frac{0.035[1.31]^3s(1+3.33s)}{(s+0.035)[(s+1.138)^2+(0.65)^2]}$$

The Laplace transform expression for the response to a unit step input would then be

$$C(s)=\frac{0.060(1+3.33s)}{s(s+0.035)[(s+1.138)^2+(0.65)^2]}$$

and its inverse time response would be

$$C(t)=K+K_1e^{-0.035t}+K_2e^{-1.138t}\times \sin(0.65t+\psi)$$

where

$$K=\frac{0.060}{0.035(1.138^2+0.65^2)}=1.0$$

$$K_1=\frac{0.060[(1-3.33)(0.035s)]}{-0.035(1.138-0.035)^2+(0.065)}=0.920$$

$$K_2=\frac{0.060(1+3.33)(-1.138+j0.65)}{0.65[-1.138+0.035+j0.65]\times[-1.138+j0.65]}$$

$$\psi=158.5 \text{ deg}$$

$$C(t)=1.0-0.920e^{-0.035t}+0.190e^{-1.138t}\times \sin\left(\frac{0.65\pi}{180}t-158.5 \text{ deg}\right)$$

Example

The block diagram of Fig. 13 describes a system. Find its closed-loop response.

Before plotting the major open loop, it will be necessary to obtain the transfer function of the closed minor loop whose forward function is

$$\frac{5}{(1+s3)(1+s10)}$$

and feedback function is

$$\frac{6.67s}{(1+s10)}$$

Fig. 14 shows a Bode plot of minor open loop

$$\frac{33.3s}{(1+s3.33)(1+s10)(1+s10)}$$

from which we may find the roots of the closed minor loop.

By inspection it can be seen that the only region that will give real root is the upward slope from $\alpha=0$ to $\alpha=0.1$. A plot of the magnitude curve in this region, obtained by adding the corrections because the two breaks at $\alpha=0.1$ and the break at $\alpha=0.3$, quickly locates this root as $\alpha=0.0186$.

The other crossover of interest is in the vicinity of 0.4 radian where complex roots will occur. The magnitude straight-line approximation has a crossover at $\omega=0.32$ radian. At this point, for $\zeta=0.60$, the lagging phase angle is found with the help of phase angle rules of Fig. 5 to add to 160.1 deg as follows

−126.87 deg	from s term in the numerator
110.00 deg	from $(1+s\ 10)$ term in the denominator
110.00 deg	from $(1+s\ 10)$ term in the denominator
67.00 deg	from $(1+s\ 3.33)$ term in the denominator
160.13 deg	

This angle is close to 180 deg so that another reading should be taken at the correct crossover obtained by including the magnitude correction curves of Fig. 3(B). Drawing these in the region of interest, we locate the actual crossover at $\omega=0.4$ for $\zeta=0.60$. Here the angle is almost exactly 180 deg so that roots are

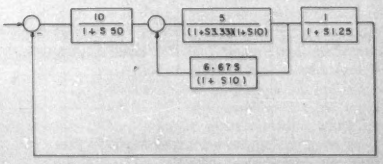


Fig. 13. Block diagram with transfer functions of a control system

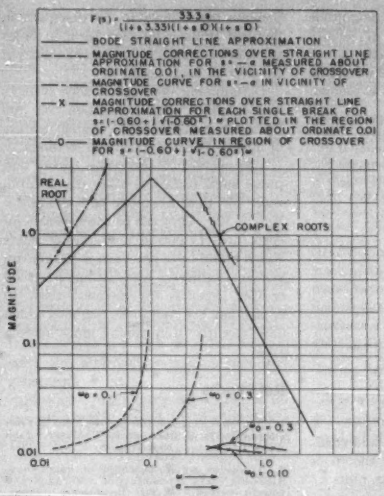
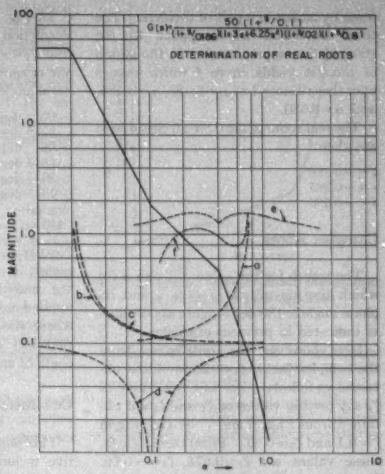

BODE STRAIGHT LINE APPROXIMATION
MAGNITUDE CORRECTIONS OVER STRAIGHT LINE APPROXIMATION FOR s = −α MEASURED ABOUT ORDINATE 0.01, IN THE VICINITY OF CROSSOVER
MAGNITUDE CURVE FOR s = −σ IN VICINITY OF CROSSOVER
MAGNITUDE CORRECTIONS OVER STRAIGHT LINE APPROXIMATION FOR EACH SINGLE BREAK FOR s = (−0.60 + j√1−0.60²)ω PLOTTED IN THE REGION OF CROSSOVER MEASURED ABOUT ORDINATE 0.01
MAGNITUDE CURVE IN REGION OF CROSSOVER FOR s = (−0.60 + j√1−0.60²)ω
REAL ROOT
COMPLEX ROOTS
MAGNITUDE
1.0
0.1
0.01
0.01
0.1
1.0
ω₀ = 0.1
ω₀ = 0.3
ω₀ = 0.3
ω₀ = 0.10
ω
σ

Fig. 14 (left). Bode plot of minor open loop function given in curve

Fig. 15 (right). Bode plot of open-loop function G(s) for $s = j\omega$ with auxiliary plot for $s = -\alpha$ for location of real roots



$(-0.60 \pm j\sqrt{1-0.60^2})0.40$

The transfer function describing the closed-loop response for the minor loop is then found as

$$\frac{5(1+s10)}{\left(1+\frac{s}{0.0186}\right)\left(1+\frac{2(0.60)}{0.40}s+\frac{s^2}{(0.40)^2}\right)}$$

i.e.,

$\alpha = 0.0186$
$\zeta = 0.60$
$\omega = 0.40$

The expression for the over-all major loop will then be

$$G(s) = \frac{50\left(1+\frac{s}{0.1}\right)}{\left(1+\frac{s}{0.0186}\right)(1+3s+6.25s^2)\times\left(1+\frac{s}{0.02}\right)\left(1+\frac{s}{0.8}\right)}$$

Real Roots

A Bode plot of this open-loop function is shown on Fig. 15.

The closed-loop response of this system can now be found by first obtaining the roots of the equation $1+G(s)=0$. An inspection of possibilities for real roots reveals that the region of 180 deg, where crossovers may occur, is in the region between $\alpha = 0.1$ and $\alpha = 0.8$. A plot of magnitude corrections for real frequencies in this region is made up of contributions from each single time constant factor as well as from the quadratic factor for which the correction curve of Fig. 3(B) applies, with $\zeta = 0.60$.

These individual corrections in the region of interest are drawn as dashed curves for convenience about the ordinate 0.1 for the single time constant terms, and about the ordinate 1.0 for the quadratic factor. Curves *a*, *b*, *c* are correction factors for terms $1+s/0.8$, $1+s/0.0186$ and $1+s/0.1$ in the denominator. Curves *d* are correction factors for the term $1+s/0.1$ in the numerator, and curves *e* are correction curves for the quadratic factor $(1+3s+6.25s^2)$. Curves *a*, *b*, *c*, and *d* are to be measured with respect to the ordinate 0.1. The difference between curve *e* and the ordinate 1 gives the corrections for the quadratic factor.

The addition of all of the contributing correction factors can be conveniently

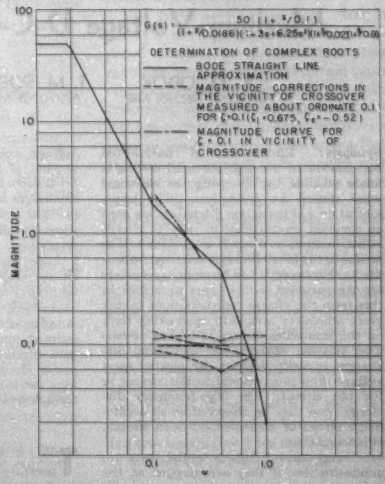


Fig. 16. Bode plot of open-loop function G(s) for $s = j\omega$ with auxiliary lot for $s = (-\zeta + j\sqrt{1-\zeta^2})\omega$ for location of complex roots. Trial value of ζ is $s = 0.1$

done by means of a pair of dividers. The total correction when applied to the straight line approximation in the region of interest yields curve f which crosses over the ordinate 1 at $\alpha = 0.186$, $\alpha = 0.385$, and $\alpha = 0.680$.

The real roots of the over-all closed loop are then:

$s = -0.186$
$s = -0.385$
$s = -0.680$

Complex Roots

The search for the value of ζ and ω for which the magnitude of G is 1 and its phase angle is 180 deg is again undertaken as indicated in previous examples.

Here, however, the contribution from the complex factor $\zeta_0 = 0.6$, $\omega_0 = 0.1$ is obtained as follows. For any trial value of ζ, say $\zeta = 0.1$, values of ζ_1 and ζ_2 are obtained from Figs. 6 and 7, by entering in $\zeta = 0.1$ and $\zeta_0 = 0.60$. From these figures, these values are $\zeta_1 = 0.675$, $\zeta_2 = -0.52$. Treat the quadratic factor as two ordinary breaks at $\omega_0 = 0.40$ but use $\zeta = 0.675$ on one and $\zeta = -0.52$ on the other instead of $\zeta = 0.1$. As ζ_2 is negative, the magnitude curves of Fig. 8 apply as well as the phase values of Fig. 9.

In Fig. 16, entering in the magnitude correction curves for $\zeta = 0.1$ ($\zeta_1 = 0.675$ and $\zeta_2 = -0.52$ for the quadratic factor) the crossover is found at $\omega = 0.2$ where the total angle is 175 deg, made up as follows:

90	deg from the break frequency 0.0186
90.5	deg from the break frequency 0.02
−68	deg from the break frequency 0.10
14.5	deg from the break frequency 0.80
29.5	deg from quadratic factor $\zeta_1 = 0.675$
18.5	deg from quadratic factor $\zeta_2 = -0.52$
175	

Next try for $\zeta = 0.20$. (For the quadratic factor, $\zeta_1 = 0.75$ and $\zeta_2 = -0.43$.) Here the magnitude correction curves when applied indicate a crossover of 0.24 radian where the angle is 191 deg. Interpolation will yield a root at approximately $\omega = 0.22$ and $\zeta = 0.14$.

Conclusions

This paper presents a simple and effective technique of determining the roots of a closed-loop system or the roots of the summation of factored transfer functions. Determination of such roots is valuable in synthesis of systems made up of minor loops or components in summation. The combinations of such components often have to be described in factored transfer-function form in order to permit a frequency response analysis of the over-all major loop.

Determining the roots of such systems also permits the evaluation of the transient response of the system by use of inverse Laplace transform methods.

The magnitude correction curves and phase angle rules which are the essential tools for this technique are shown on Figs. 5–10. They can be conveniently incorporated in plastic templates or rulers which would be of great use to the control systems engineer. The method described in this paper is considerably easier to apply than the root locus methods which have recently been gaining much acceptance.

References

1. Servomechanisms and Regulating System Design (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., New York, N. Y., vol. I, 1951.

2. Discussion by A. Leonhard of Frequency-Response Data Presentation Standards and Design Criteria, Rufus Oldenburger. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 76, no. 8, Nov. 1954, pp. 1155–76.

3. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

Improved D-C High-Potential Testing of Insulation Systems in Low- and Medium-Voltage D-C Equipment

A. M. ODOK
ASSOCIATE MEMBER AIEE

T. M. SOELAIMAN
ASSOCIATE MEMBER AIEE

Synopsis: Elimination of unexpected breakdowns in d-c high-potential testing is made possible by observing the electrical noise caused by ionization. (The terms ionization and ionization discharges are used as common terminology to refer to charge redistributions as a result of gaseous breakdowns in the insulation systems; see Appendix I. However, it will be more appropriate for the purposes of this paper to include in this term the charge redistributions taking place whenever a minute conducting path on the insulation surface is broken down.) The relative severity of ionization discharges is easily detected by picking up a voltage drop proportional to the current in the test circuit. The high-frequency discharges can then be observed on an oscilloscope screen or measured by other means. It is found that:

1. Direct-current high-potential tests are nondestructive if they are stopped at the voltage value at which any appreciable amount of ionization is observed, because breakdown is always preceded by ionization discharges in the insulation system.
2. The noise pattern on the oscilloscope screen is different for different conditions of insulation. This, in conjunction with the leakage- and absorption-current characteristics, enables one to tell whether the insulation is dry or wet, clean or dirty.

Tests have been made on small laboratory samples of insulation; on components of railroad equipment, such as armatures and field windings; and on entire locomotive power circuits. A great number of tests were carried out without any unanticipated breakdowns occurring.

TWO METHODS are at present widely used for evaluation of insulation systems in machines and equipment: the low direct-voltage resistance measurement (so-called megohmmeter tests); and the a-c high-voltage tests with prescribed overvoltages and measurement of the loss angle.

Both of these methods are subject to criticism, and much research and development work is still being done to find a satisfactory substitute. The main reasons for dissatisfaction with the test methods include the fact that the megohmmeter readings are not conclusive for comprehensive evaluation of insulation. Machine insulations with a very low insulation resistance, as measured with an ohmmeter, are very often known to give satisfactory service. An example of this is the electric equipment of diesel-electric locomotives in very humid atmospheres. If the low megohmmeter reading were to be taken as a measure, many of the locomotives would be taken out of service.

Paper 59-247, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 28, 1958; made available for printing January 28, 1959.

A. M. Odok is with the General Electric Company, Erie, Pa., and T. M. Soelaiman is with Universitas Indonesia, Djakarta, Indonesia.

However, experience shows that most of these locomotives operate satisfactorily. Alternating high-potential tests, on the other hand, which include loss-angle measurements, are complicated and require bulky equipment because of the large reactive power needed to charge the machine capacitances. Apart from this drawback, the tests are mostly of the go-or-no-go type and are based on fixed overstress determined by experience.

Mainly for the afore-mentioned reasons, various scientists and engineers have devoted much time to investigating the merits of high-potential direct current for evaluation of insulation systems.[1-16] The main issue in the previous work has been the validity and methods of interpretation of the d-c high-potential test results for high-voltage a-c motors and generators.

Since d-c high-potential testing is inherently better suited for d-c machines and equipment than for a-c machines, some of the controversies arising in its use on a-c machines are of less significance here. However, the d-c high-potential test procedure proposed and discussed in various publications to date is still not directly useful even for d-c equipment because of certain oversimplifications in interpreting the test results, particularly those concerning the physical picture of insulation behavior.

In this paper it will be shown that the d-c high-potential tests, when supplemented by observation of ionization discharges during test and when properly interpreted, can provide information about the insulation condition. Further, a test procedure will be described, and results obtained on specimens as well as on d-c machine components will be discussed. Finally, some results of testing the insulation of entire power circuits of locomotives will be presented.

Discussion

Aim of D-C High-Potential Testing

In the most general terms, the aim of an insulation-system test is to predict whether the system will withstand, for a prescribed period of time, the dielectric stresses imposed by its mode of operation. Therefore, the normal operating stresses and the nature of the breakdown must be considered when designing an insulation test method and interpreting experimental results.

The dielectric strength of insulation is affected by one or more of several conditions: mechanical damage; chemical deterioration; and the presence of dirt, dust, and moisture. The mechanism of breakdown varies with the insulation condition; basically it can be either intrinsic or result from discharges and thermal instability (see Appendix I).

Since this study is restricted to low- and medium-voltage d-c equipment, the expected operating stresses and mechanism of breakdown may be defined as follows:

1. Stresses. By low and medium voltage is meant those operating voltages for which test voltages do not exceed 8 kv. The operating stresses are supposed to be constant. (This is actually true of the field windings. The armature, however, is subjected to stresses which vary from zero to a maximum. These correspond to superposition of d-c on a-c stresses.)

2. Breakdown mechanism. Ionization discharges and resulting thermal instability are the most probable mechanisms involved in the breakdown of insulation systems considered here.

On the basis of these definitions and assumptions, it follows that the test method should be designed to satisfy the following conditions:

1. Insulation system under test should be stressed by direct voltage.

2. Any discharges which may induce the breakdown of the insulation should be detected before reaching any dangerous extent.

3. The cause for a low ionization starting voltage should be detected reliably.

In other words, the purpose of the test method is to determine the highest voltage level at which the given insulation will be free of damaging ionization; which, presumably, is the first stage in the breakdown of insulation systems under consideration here. Moreover, the results should show whether discharges are attributable to moisture, dirt, or conducting paths on creepage surfaces.

Shortcomings of Present Techniques

The present d-c high-potential test procedure and methods of interpreting its results have been described in numerous papers.[1-16] This discussion of their limitations is with a view toward improving the method of interpretation. It avoids a controversial part of the previous work, namely, whether information gained under d-c stresses can be of any use for a-c applications and, specifically, whether the connection between the breakdown voltages under the two kinds of stresses can be expressed as a single number, such as the 1.6 d-c–a-c ratio adopted by some investigators.

The present technique of d-c high-potential testing consists of measuring the following two quantities:

1. The steady-state leakage current is measured as a function of voltage applied between conductors and ground (leakage-current characteristic).

2. The leakage current is measured as a function of time with voltage held constant (absorption- or polarization-current characteristic).

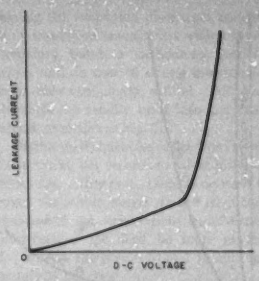


Fig. 1. Typical leakage-current characteristic for insulation in poor condition

In all previous work the evaluation of insulation has been based on these quantities, with a few exceptions where ionization probes have been used to detect electric discharges at certain locations of the winding.[6] Foust and Bhimani have used the additional tool of detecting the over-all electrical noise.[11] As a consequence of the limited information gained in this way, the publications on the subject differ mainly in the interpretation of these two measured quantities and so a brief discussion of the proposed ways of interpreting the leakage-current characteristic is appropriate.

The curve $i=f(v)$ has been used in different ways for predicting the breakdown voltages: In all cases where polar materials, such as water and certain types of contaminants, are present or a "deteriorated" insulation is being tested, the leakage-current characteristic will have a more or less pronounced knee after which the current increases more rapidly with increase in voltage (Fig. 1). This led to the hope that an asymptote may be found to indicate the value of voltage at which the current would increase to infinity. This voltage would then be looked upon as a breakdown level. The arbitrariness in this procedure can be demonstrated by changing the scale for the current ordinate. Evidently the geometrical shape of the leakage-current curve and the location of the visual asymptote are dependent on the scale chosen. Moreover, if this method is used with a good, dry insulation, it will lead to breakdown during test because no knee will be observed to indicate the approaching failure. With

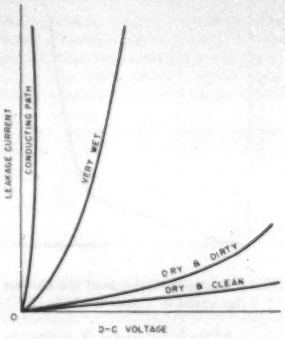


Fig. 2. Typical leakage-current characteristics for various conditions of insulation

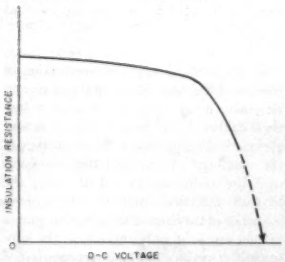


Fig. 3. Typical insulation resistance characteristic

changing amounts of humidity almost any shape of curve can be obtained, ranging from a very rapid increase of current with very wet insulations to the low-current straight-line characteristic of good, dry insulation (Fig. 2).

Another version of the foregoing method of prediction has been proposed in which the resistance characteristic $R=f(V)$ is extrapolated to $R=0$. The voltage at which the resistance approaches zero is taken to be the breakdown voltage (Fig. 3).

The leakage-current characteristic alone cannot tell anything about the presence of polar materials, such as moisture or contamination, because rapid increase of current is a relative condition which is difficult to recognize unless the characteristic of the clean, dry insulation for the machine under investigation is available. In the absence of this information, the absorption-current characteristic must be used to determine whether the insulation is contaminated or moist.

Fig. 4 (right). Simple circuit for ionization detection

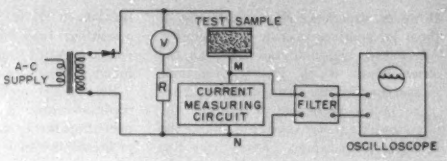
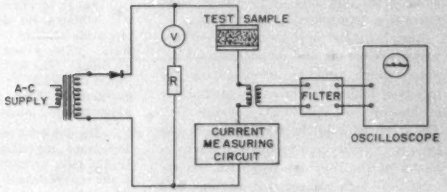


Fig. 5 (right). Ionization detection circuit with insulating transformer



Improved Test Method

An observation made by the authors, and substantiated by careful study of the previously published work, can be made the basis of a nondestructive test method.

None of the insulation systems investigated broke down without discharges starting at a lower voltage level than the ultimate breakdown value. The discharges were external, internal, or on the surface, but usually two or all three kinds occurred simultaneously. (See Appendix I for explanation of terms.) Consequently, an ionization detection method which would cover all three kinds of discharges in the insulation system under test would indicate any approaching breakdown. Figs. 4 and 5 show two possible arrangements for this purpose. Essentially, this circuit consists of a means for picking up the voltage drop across the current-measuring circuit or an inductor coil, preferably through an insulation transformer, and an oscilloscope for observation of this voltage. Filters may be used to eliminate external noise resulting from industry frequency and other possible sources. The simple *R-C* high-pass filter has been found quite satisfactory. If it were desired to detect only the start of a certain ionization level some type of warning device could be used instead of the oscilloscope. However, the discharge pattern on the oscilloscope supplies additional valuable information (see Fig. 6).

Discharges start in a moist, contaminated insulation system at a much lower voltage level than the actual breakdown voltage. The test therefore usually will be stopped long before a breakdown can be anticipated. This may result in overcautiousness so that the anticipation of breakdown will be too far on the safe side. This circumstance, which can to a great extent be corrected by experience and skill in testing, is no drawback be-

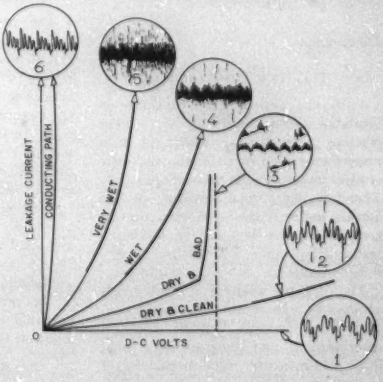


Fig. 6. Typical oscilloscope patterns for various conditions of insulation

1—No discharges
2—Occasional discharges
3—Broken down
4—Frequent discharges
5—Hazy picture due to very high frequency of discharges
6—Conduction only, no discharges

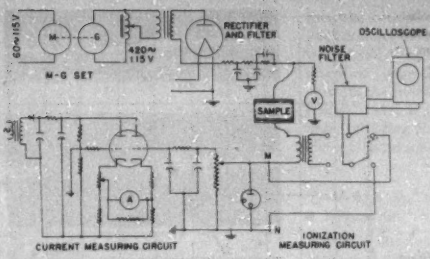


Fig. 7. Schematic of main circuits

cause moist and contaminated insulations will usually be treated in some way to remove polar materials before being put into service and subjected to normal operating stresses.

The cleaner and drier the insulation, the more closely the discharge starting voltage approaches the actual breakdown value. Therefore, the prediction of the breakdown voltage in clean and dry insulation without conducting paths on the surface is rather accurate. With the start of discharges of any appreciable magnitude and frequency the test must be stopped because the breakdown follows these discharges closely.

When the observation of discharges is used together with absorption-current versus time and leakage-current versus voltage characteristics, it will reliably indicate:

1. The presence of polar materials and conducting paths (moisture, dust, dirt, carbon tracks) and gross insulation defects.

2. The maximum direct voltage which can be impressed on the insulation system without damaging it.

3. The breakdown voltage under d-c stresses for a clean and dry insulation. This value contains a certain safety margin because breakdown occurs at a higher voltage than that at which discharges start.

Certain insulation weaknesses, such as turn-to-turn faults, mechanical damage, and delamination, cannot be reliably detected by a d-c high-potential test unless they have become sources of ionization previous to the time of tests. However, since no other single high-voltage test method will reveal all of these inherent weaknesses, this circumstance cannot be looked upon as a drawback to the proposed technique.

It should be noted that the observation of electrical noise need not necessarily be restricted to d-c stresses, but may perhaps be used with a-c stresses as well. This would make a-c high-potential tests nondestructive also. It is conceivable that discharge detectors could be built into the three phases of a generator, making it possible to follow the insulation quality closely under operating conditions.[9]

Equipment and Test Procedure

Before presenting results obtained by this technique of d-c high-potential testing, the test equipment and procedure will be explained in some detail.

Equipment and Circuitry

The principal circuit of a d-c high-potential test set is shown in Fig. 7. If the equipment to be tested is grounded, point *M* instead of *N* can be put on ground potential.

Additional equipment needed for observing the ionization discharges consists of an insulation transformer with a broad frequency band, a cathode-ray oscilloscope, and, in case of strong interference from other equipment, a filter. All three elements are readily built into the basic circuit, as shown in Fig. 7.

Discharge frequencies range from about 1,000 cycles per second to a few hundred kilocycles. However, there is no need to cover the whole range, so the frequency response of the series circuit transformer-filter-oscilloscope need be more or less flat only in the region from 1 to approximately 50 kc.

Test Procedure

The d-c high-potential test procedure involving only the absorption-current versus time and the leakage-current versus voltage measurements has been repeatedly described and discussed.[1–10] Although this paper is principally concerned with the technique and advantages of studying ionization discharges, certain observed peculiarities of absorption and leakage current characteristics will also be considered.

Absorption-Current Characteristics

When a constant direct voltage is suddenly applied to an insulation system, a rather high current inrush is observed. After a certain time the current settles to a lower steady value called the leakage current (Fig. 8). This time-dependent absorption current contains several components with different time constants. The longest time constant is due to a physical process called interfacial or space-charge polarization. The high sensitivity of this component to the presence of water or other polar material makes the ratio i_1/i_{10} (the absorption coefficient) meaningful for insulation evaluation. Here i_1 is the current 1 minute and i_{10} the current 10 minutes after application of the test voltage (Appendix II).

The presence of polar molecules causes

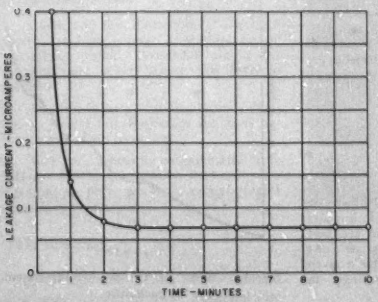


Fig. 8 (left). Typical dielectric - absorption characteristic of dry insulation in good condition

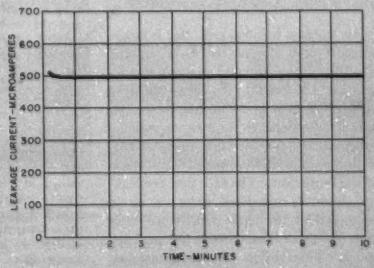


Fig. 9 (right). Typical dielectric - absorption characteristic of wet insulation in poor condition

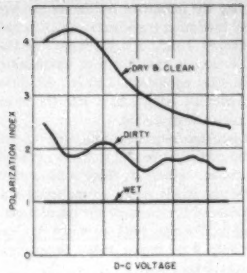


Fig. 10 (left). Typical polarization indexes for various conditions of insulation

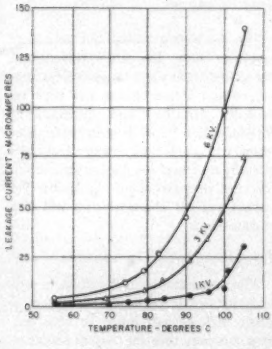


Fig. 11 (lower left). Effect of temperature on leakage-current characteristics

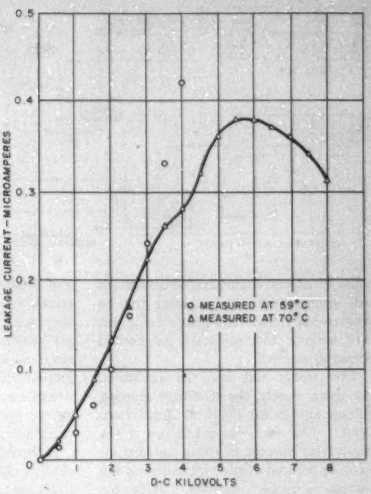


Fig. 12 (right). Effect of humidity on leakage-current characteristic

insulation to behave more like a pure ohmic resistance. This is characterized by a very short time constant for the establishment of the current (Fig. 9).

In choosing the voltage level for the absorption test the following two facts must be considered:

1. The polarization index depends on the voltage at which the absorption-current characteristic is being measured (Fig. 10).

2. For humid or heavily contaminated insulation, a low test voltage must be chosen for the absorption-current characteristic since discharges start very early and can be damaging to insulations under these conditions (Fig. 6).

Leakage-Current Characteristic

As already pointed out, the leakage-current characteristic in itself supplies practically no information about the insulation condition. Its shape is only indirectly related to discharge severity and its geometrical slope is as much a function of the chosen scale of co-ordinates as it is of the physical condition of the insulation. The fact that the leakage current is considered under a separate heading does not imply that its use apart from the absorption-current test and discharge observation is recommended. When supplemented with these two, however, it serves a certain purpose.

For practical reasons, the applied voltage is increased in finite steps during the leakage-current test. Consequently, with each increase of voltage some time must be allowed to establish a steady-state current. For the kind and size of machines here considered and the voltage steps used (500 to 2,000 volts), experience shows that 1 minute at each voltage step is long enough for the current to settle at a practically constant value.

If the preceding absorption-current test has shown the insulation system to be moist or dirty, the initial voltage must be very low, small voltage steps must be used, and electric discharges must be observed closely.

Some peculiarities of the leakage-current characteristics found by experimental analysis:

1. Shape of the curve. If the same insulation is tested under different conditions and the leakage-current characteristics are plotted using the same scale for the co-ordinates, certain trends are observed depending on moisture and dirt content and the presence of conducting paths in the insulation system. A schematic representation of various curve shapes is given in Fig. 2.

2. Effect of temperature. Temperature has a bearing on the magnitude of the leakage current. Results of an experiment are shown in Fig. 11. From this it appears that tests at room temperature are best suited for purposes of comparison. At these temperatures the current for a given voltage is least affected by changes in temperature.

3. Effect of repeated tests. A repetition of test usually gives a different leakage-current characteristic. The current tends to decrease with repetition in the case of clean dry insulations, and to increase if the insulation is moist. In this respect, dirty-dry insulations usually behave like clean-dry insulations and dirty-wet ones like moist insulations.

4. Effect of humidity. Under certain circumstances humidity can give rise to a characteristic such as shown in Fig. 12.

5. Effect of dirt and conducting paths. Sometimes no increase in leakage current with increasing voltage is observed until a certain voltage level is reached. Above this the current starts to increase until it reaches another steady value at a higher voltage and so on. The leakage-current characteristic in this case looks as shown in Fig. 13. A possible explanation of this phenomenon is given in Appendix III.

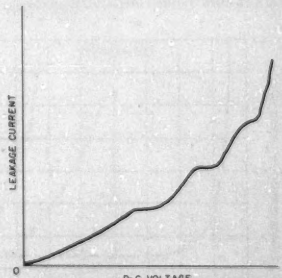


Fig. 13. Effect of dirt on leakage-current characteristic

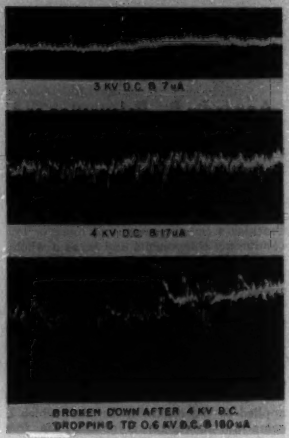


Fig. 14. Oscilloscope patterns for insulation samples (moistured mica mat, double sheets)

Ionization Discharge Characteristics

The oscilloscope screen must be observed during the recording of the leakage current versus voltage to determine the discharge pattern, since discharges are a measure of the damaging effect of the test to the insulation.

Occasionally discharges will occur at relatively low voltages. Persistent and severe discharges are a sign that damage to the insulation is starting. Tests must be stopped at the first sight of these discharges, unless breakdown of the insulation is intended.

Expected discharge patterns under various conditions of the insulation are shown in Fig. 6. These assume that the usual interference from industry frequency is not completely filtered out.

Following are some characteristics of the discharge patterns, depending on insulation condition:

1. Clean and dry insulation. Discharges start at a high voltage level close to the breakdown value. Occasional single discharges may occur before this level is reached.

2. Dry but slightly dirty insulation. Similar to item 1 but discharges are more frequent and easier to detect.

3. Wet and/or heavily contaminated insulation. Discharges start at very low voltages and the pattern is hazy, a sign of higher frequency discharges.

Test Results

In the following section some ionization discharge characteristics obtained by using the methods described are discussed. Furthermore, tests on the insulation systems of two diesel-electric locomotives are described. Since the purpose was to demonstrate the usefulness of ionization discharge study, a thorough discussion of tests, including absorption- and leakage-current characteristics, seemed unnecessary.

Tests on a Small Sample of Insulation

The oscillograms in Fig. 14 show the increase of discharge frequency and magnitude with increasing voltage on a double-sheet sample of moistened mica insulation. Discharges started at approximately 3.5 kv and increased rapidly. The pattern at 4 kv indicated that the discharges had assumed a magnitude damaging for the insulation. The sample broke down as the voltage was being increased to the next step of 4.5 kv. Because of the high internal resistance of the test set, the voltage dropped to a low value after breakdown.

Tests on Machine Parts

Each of the oscillograms reproduced in Figs. 15(A) and 15(B) shows two curves corresponding to the two methods (Figs. 4 and 5) used for picking up the voltage drop proportional to the discharge current. The curve with high ripple content is obtained by using the arrangement shown in Fig. 4 and the smooth curve by using a transformer shown in Fig. 5. The transformer used attenuated the lower frequencies, thus eliminating the ripple on the d-c high voltage.

1. Test of a d-c motor field coil. Fig. 15(A) shows the discharge characteristics of a field coil which was dirty but dry. Discharges started at 9 kv, and the insulation broke down at the next voltage step.

2. Test of a d-c motor armature. The oscillograms of Fig. 15(B) were taken on a motor armature which was wet and dirty. The noise detected at lower voltages by the transformer pickup is external. The insulation broke down as 2 kv was exceeded. Note the hazy discharge pattern indicating the presence of moisture and dirt.

Tests on Locomotive Power Circuits

The insulation systems of two diesel-electric locomotives were tested. Although the ionization discharges were observed on the oscilloscope, no oscillograms were taken. Only a summary of observations made during the test will be given since these tests are yet in preliminary stage.

Tests followed the usual sequence of (1) a visual inspection of the insulation, (2) the dielectric absorption test, and (3) the

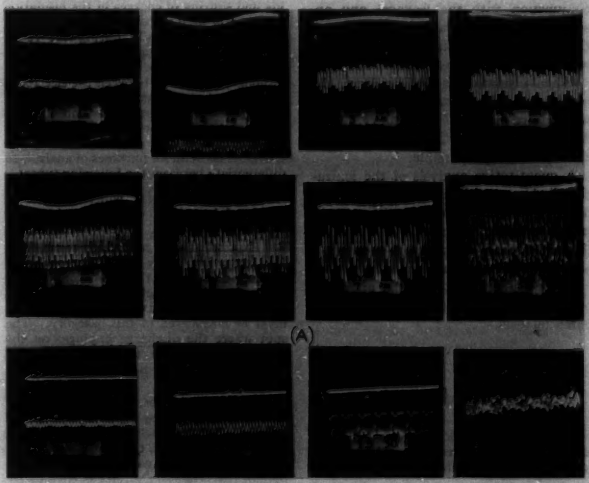


Fig. 15. Oscilloscope patterns obtained with test circuits shown in Figs. 4 and 5

A—D-c motor field coil
B—D-c motor armature

leakage-current test, together with the observation of ionization discharges.

In each case, the circuit under test consisted of one generator, two motors, and the power supply cable.

The inspection revealed that all components accessible to visual inspection were in a highly contaminated state.

The polarization index was practically unity, and the leakage current increased very rapidly, reaching the maximum output of the test set (2-milliamperes) at 350 volts. Discharges started at a low voltage and the discharge pattern was hazy.

A diagnosis, based on the electrical testing only, would have reliably detected what in this special case could be found by visual inspection. Further tests on individual components showed that the motors were free of ionization discharges up to 1,000 volts and had a low leakage current. The generators and cables were found to be the main source of high leakage and excessive discharge currents.

Conclusions

If, in addition to the usual measurement of absorption and leakage current, means of electric discharge detection are employed, d-c high-potential tests yield additional information and become nondestructive.

Observation of ionization discharges, in conjunction with the leakage- and absorption-current characteristics, indicates the condition of the insulation, i.e., whether it is dry or wet, clean or dirty.

When insulation is wet or dirty, the breakdown voltage can be predicted only with a large margin of uncertainty. With dry, clean insulation, however, a much closer prediction is possible.

A quantitative analysis of the ionization discharges with respect to magnitude, power, and frequency spectrum may supply valuable information about insulation condition. This aspect presents a promising field for further investigation.

It is suggested that a similar approach be investigated for use during a-c high-potential tests.

Appendix I. Theories of Breakdown

The following three kinds of breakdown of solids are known to occur either singly or in combination under dielectric stresses.[12-18]

Intrinsic Breakdown

This is best defined as the type of breakdown which occurs at very high voltages if the possibility of the two other kinds of breakdown is well excluded. Its mechanism is almost certain to be electronic (theories by von Hippel and Fröhlich). The intrinsic dielectric strength is a function of time, temperature, and sample thickness (for very thin samples), and its value is of the order of magnitude 10^6 to 10^7 volts per centimeter.

Thermal Breakdown

The heat generated in conducting paths and/or the dielectric losses with a-c stresses change with temperature as well as with voltage. Hence a thermal instability can occur in which the generated losses cannot be dissipated as rapidly as they are produced. Consequently, very high local temperatures can occur, leading to destruction of the insulation material.

This type of breakdown is usually the second stage in a breakdown caused by d-c stresses, the first phase being some kind of discharge impinging on the insulation surface.

Breakdown by Discharges

The test sample or machine consists of three elements: the two electrodes and a composite insulation. The latter is not of a homogeneous structure, and may have many more or less conducting cross paths inside and on the surface, and voids of different sizes. This system is best regarded as a random lattice network of capacitors and resistors. For a given potential difference between the electrodes there is a certain potential and current distribution in this lattice. If, however, the potential difference between any two points exceeds the ionization value for the gas between these points, the gas will be conducting and shifting of charges in the insulation system will take place. With the new distribution, or especially during the redistribution, some new gaps will be conducting. Therefore, random discharges will occur, although the voltage between the electrodes may be kept at a constant value.

Despite the fact that ionization of a gas adjacent to or trapped-in the insulation is the common feature for all types of discharge, it is useful for purposes of analysis to differentiate between the following types:

1. Internal discharges, which result from the breakdown of the gases in the voids inside the insulation.

2. External discharges, which result from the formation of an ionization path between the surface of the insulation and one electrode. Any discharges between the slot wall and wall insulation in a machine must be considered in this category. The spot where the discharge impinges on the insulation surface is usually the starting point of a breakdown caused by thermal instability.

3. Surface discharges, which is most likely to occur in machines, especially under d-c stresses. Surface discharges occur between conducting spots on the insulation where high enough stresses can build up if a large part of the insulation surface is covered by such conducting paths or substances as carbon tracks, water, and dirt. The discharge develops along the surface between the conducting spots. After each discharge a redistribution of potential on the surface takes place. This phenomenon keeps repeating, so that a rather-high-frequency current component is observed in the external circuit.

Appendix II. Polarization and Absorption

Maxwell's Theory on Composite Dielectrics

Any insulation material used in practice contains a certain amount of more or less conducting components and is, as a whole, nonhomogeneous with voids of different sizes. A composite insulation has the additional feature of consisting of layers of materials differing in their dielectric as well as their conducting properties. Thus, electrically speaking, a composite insulation can be considered as representing a lattice network of randomly connected resistors and capacitors of various sizes.

The instant a constant direct voltage is applied to the electrodes between which a composite insulation is placed, the free charges (and to a certain extent the bound charges) will start to move. Dielectric dipoles will orient themselves according to the applied potential difference between the electrodes until a balance between the applied voltage and the back potential is established. This back potential results from the charge distribution and the voltage drop caused by the current across the conducting paths. Because of this the initial current will contain a part attributable to charge displacements and dipole orientation in addition to a part flowing through the conducting paths. The final current will be only the conduction current flowing through paths which extend from one electrode to the other.

Polarization in Homogeneous Materials[14,15]

Three principal kinds of polarization are possible in a homogeneous solid insulation:

1. Polarization caused by induced dipole moments. Electrons surrounding the atomic nuclei or unequally shared by atoms of a molecule will tend to follow a distorted trajectory under the influence of an external field, thus giving rise to an increase of capacity. The so-called electronic and atomic polarizations are of this type.
2. Orientation polarization is attributable to the fact that an external field will tend to orient the dipoles already existing in the insulation in the absence of this field.
3. Space charge or interfacial polarization are caused by charges which can move inside of the dielectric until they become trapped in the material or on interfaces.

Appendix III. Effect of Discharges on the Shape of the Leakage Current Characteristic

Experimental studies imply a definite relationship between the increasing slope of

the leakage current characteristic and the frequency of electric discharges. The following hypothesis about the mechanism behind the shape of the leakage current characteristic may help to explain some of the peculiarities observed experimentally.

The picture of a random lattice network of resistances and capacitances underlying Maxwell's theory of composite insulations must be supplemented, in the case of actual insulation, by certain assumptions about the potential and current distribution on the surface of the insulation, e.g., on the surface of the end windings. This surface is far from being absolutely clean; in addition to water, it carries dust and certain other contaminants. Carbon deposits will also be found on this surface, especially in cases of machines with carbon brushes or organic insulation.

Aggregates of such conducting or polarized materials are at practically equal potential. Consequently, the total potential difference is distributed only on the small remaining area which is free of such foreign matter.

As explained in Appendix II for the interfaces and inside of the insulation, the potential distribution on the surface will be disturbed if the field strength at any point exceeds the ionization level of the adjacent air. The redistribution of charges on the surface leads to a new balance of potentials. A certain leakage current corresponds to the newly established conducting paths. With increasing voltage, rearrangement of potentials inside and on the surface of the insulation will occur. The more discharges occur at a given voltage, the higher the stress imposed on the remaining parts of the insulation so that with the beginning of any appreciable amount of discharges an exponential take-off of the current would be expected. In fact this is observed in all insulation systems where the surfaces are contaminated or moist.

With dry insulations, the leakage current characteristic more closely follows a straight line. The insulation behaves like an ohmic resistor and no shifting of charges takes place as voltage is increased. Of course, this is true only up to that voltage at which even the more uniformly distributed potential of the dry insulation will be enough to start ionization. A more sudden increase of ionization and insulation damage is to be expected in this case since, with the disturbed balance of potential distribution, other points already stressed to the very limit will be overstressed and break down.

The step-like leakage current characteristic often observed with dirty insulation (Fig. 13) can be explained on the basis of discharges on the surface. During the increase of voltage, a point in the potential distribution on the surface may be reached above which a further increase of voltage produces no new discharges. The current may then stay at a practically constant value until the voltage is increased to a value high enough to create new ionization channels.

References

1. Dielectric Strength Ratio Between Alternating and Direct Voltages, J. L. R. Hayden, W. N. Eddy. *General Electric Review*, Schenectady, N. Y., vol. 26, Sept. 1923, pp. 646–52.

2. Maintenance Over-Potential Tests for Armature Windings in Service, R. N. Wiseman. *Ibid.*, vol. 53, Aug. 1950, pp. 24–28.

3. Alternating and Direct Voltage Endurance Studies on Mica Insulation for Electric Machinery, G. L. Moses. *AIEE Transactions*, pt. I, vol. 70, 1951, pp. 763–69.

4. Nondestructive Testing of Generator Insulation, E. H. Povey, F. S. Oliver. *Electrical Engineering*, vol. 70, June 1951, p. 498.

5. A Maintenance Inspection Program for Large Rotating Machines, J. S. Johnson. *AIEE Transactions*, pt. I, vol. 70, 1951, pp. 749–55.

6. Diagnosis of A-C Generator Insulation Condition by Nondestructive Tests, A. W. W. Cameron. *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 71, Jan. 1952, pp. 263–74.

7. Techniques and Examples of High-Voltage D-C Testing of Rotating-Machine Windings, C. L. Sidway, B. R. Loxley. *Ibid.*, vol. 72, Dec. 1953, pp. 1121–29.

8. Leakage-Voltage Characteristics of Insulation Related to D-C Dielectric Strength, J. S. Johnson, J. W. Clokey. *Ibid.*, Aug. 1953, pp. 681–86.

9. Examination of High-Voltage D-C Testing Applied to Large Stator Windings, R. T. Rushall, J. S. Simons. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 102, pt. A, Oct. 1955, pp. 665–80.

10. D-C High-Potential Maintenance Testing of Traction Motors and Generators, W. Schneider. *AIEE CP56-372* (available on request).

11. Predicting Insulation Failures with Direct Voltage, C. M. Foust, B. V. Bhimani. *AIEE Transactions*, pt. III, vol. 76, Dec. 1957, pp. 1120–39.

12. The Insulation of Electric Equipment (book), W. Jackson. John Wiley & Sons, Inc., New York, N. Y., 1954.

13. Dielectric Breakdown of Solids (book), S. Whitehead. Oxford Clarendon Press, New York, N. Y., 1951.

14. Dielectrics and Waves (book), A. V. Hippel. John Wiley & Sons, Inc., 1954.

15. Dielectric Materials and Applications (book), A. V. Hippel. John Wiley & Sons, Inc., 1954.

16. Some Recent Studies of Random Noise Between Metals and Dielectrics, S. I. Reynolds. *Journal of Applied Physics*, New York, N. Y., vol. 27, 1956, pp. 728–34.

Discussion

A. Wichmann (Siemens-Schuckert-Werke AG, Mülheim, Germany): The paper is a very useful contribution to the development of d-c insulation test methods and a better understanding and analysis of d-c ionization measurements. It would be of interest to know what method was used to distinguish between external, internal, and surface discharges, and possibly disturbances from outside. In our very similar investigations on high-voltage insulation for turbogenerators made some years ago in our laboratory it was difficult to avoid disturbances from outside of the test specimen. However, the test voltage needed for internal ionization was higher than the voltage used by the authors.

We tried to separate the internal ionization from discharges on surface and outside in a special way.[1] In this method the test bar insulation is charged with direct current for a certain period to obtain good polarization of the insulating material, and after ionization in the holes a nearly complete compensation of the charge on the hole by charges of other polarity on the surface of the holes as shown in Fig. 16(A). Removing direct voltage and connecting both electrodes, the polarization of the solid insulation is equalized in a time given by the absorption current equation.

$$i_a = Kt^{-n}$$

The surface charges in the holes are no longer compensated for and thus cause increasing voltage and, finally, a gas breakdown to equalize the positive and negative charges. These breakdowns may be seen as current pulses on an oscilloscope screen or be counted by an impulse recorder. The pulses are now of opposite polarity as during charging time. The number of pulses per second is decreasing with time and increasing with voltage. The slope of the curve given in Fig. 17 is very similar to the tan δ increase with voltage caused by a-c glow discharges in the holes of the insulation.

The ionization caused by direct or alternating current will not act as a signal of breakdown if the insulation contains mica flakes. High-voltage insulation for electric machines contains large mica splittings and, therefore, it is able to withstand glow discharges during service or test over a very long time without damage and failure. The insulation system tested by the authors seems to be made without mica, otherwise it would not be understood that d-c ionization is followed by breakdown. Ionization in generator insulation normally starts at

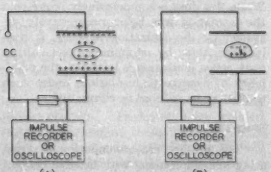


Fig. 16. Direct-current polarization and ionization during (A) charging and (B) discharging time

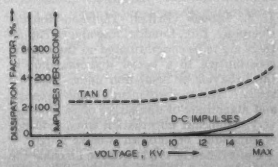


Fig. 17. Tan δ and d-c ionization pulses during first 5 seconds' discharging time on generator bar insulation

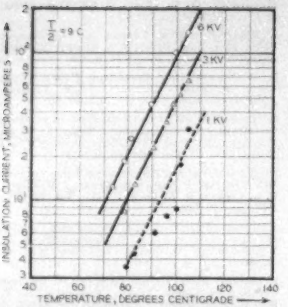


Fig. 18. Effect of temperature on leakage-current characteristic (based on Fig. 11)

rated voltage, sometimes lower, while breakdown occurs at ten times higher voltage, and very seldom at only two or three times higher voltage if the insulation is mechanically deteriorated. Therefore, the experience of the authors as to the indication of breakdown by ionization is limited to their special insulation system.

I also would like to comment on Fig. 11 of the paper which shows the effect of temperature on the leakage-current characteristic. If the points given in Fig. 11 are plotted versus voltage it may be seen the leakage current follows Ohm's law at any temperature between 55 and 105 degrees centigrade. In plotting leakage current versus temperature as shown in Fig. 18 we recognize the normal temperature dependency valid for all insulators and given by the formula

$$i(\xi) = i(\xi_0)e^{\alpha(\xi - \xi_0)}$$

Measurements at different temperatures may be reduced to the same temperature for comparison by a straight line on log paper without difficulty.[8] If we apply this knowledge to the measurements given in Fig. 12 we must conclude that the decreasing leakage current with increasing voltage may be caused by drying during test.

References

1. Maintenance Testing of High Voltage Insulation in Electrical Machines, A. Wichmann. *VDE Fachberichte*, Wuppertal-Elberfeld, Germany, vol. 19, 1956, pp. I/136–46.

2. Drying of Electrical Machines, A. Wichmann. *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 76, 1955, pp. 340–47.

J. S. Simons (British Thomson-Houston Company, Ltd., London, England): The authors are to be congratulated on their useful contribution to a better understanding of the behavior of insulation under high d-c stress. The measurement of the presence and magnitude of discharges often provides very significant information when evidence of impending breakdown is being sought.

Unfortunately the larger the machine the more difficult is the interpretation of results, as the discharges observed may represent the summation of many separate sites each consisting of a cluster, and involving external, internal, and surface types of discharge.

When we recently carried out high-voltage d-c tests both on individual stator coils and on a complete stator winding, in many cases breakdown was preceded by visible and audible discharging. However, in some cases failure occurred suddenly and in such cases even ionization detection measurements would not prevent serious damage being done before the test could be stopped. In the majority of cases ionization measurements would assist substantially in the interpretation of results.

We favor a-c testing for large a-c machines and we have carried out a-c ionization measurements both on complete windings and on individual coils. Such measurements assist materially in assessing insulation quality. The use of discharge detection methods is to be encouraged and the authors have demonstrated the potentialities of such techniques.

B. V. Bhimani (General Electric Company, Schenectady, N. Y.): The authors' conclusive evaluation of direct-voltage tests, applied to composite insulation systems of rotating machines, is a significant contribution to the field of insulation testing. There is considerable information in their paper that attracts attention, and it is specifically based on tests performed on complete machine windings.

Their evaluation separates the reliable aspects of direct-voltage tests from limitations of such tests. Their recommendation of measuring ionization, along with leakage current and polarization index of insulation, seems to increase the reliability of interpretation of such data. Another important aspect of this recommendation is that it reduces the possibility of damage to insulation during direct voltage tests. These aspects of the authors' work are discussed in the following.

Evaluation of Direct-Voltage Tests

Reliable Aspects

The authors reaffirm that direct voltage tests are reliable indicators of the following:

1. Dirt, dust, and "moisture" content of insulation.
2. Presence of carbon tracks of conducting particles in or on an insulation tested.
3. Some gross insulation defects, e.g., short circuits from conductor to ground.

They also recommend the addition of ionization measurements, to those of leakage current and polarization index, for increasing the reliability of interpretation of direct voltage test data.

Limitations

Limitations of direct-voltage tests, as indicated by the authors, are

1. The effect, on insulation strength, of mechanical damage and chemical deterioration, cannot be reliably detected by such tests.
2. These tests cannot reliably detect turn-to-turn faults in a winding, and insulation delaminations, i.e., insulation that has satisfactorily withstood such tests may still have serious "weaknesses" that could affect its operating characteristics.
3. Results of such tests are affected by the temperature of the insulation tested.
4. Leakage current readings may vary from one direct-voltage test to another, on the same insulation, even when factors such as temperature, humidity, dirt content, and "condition" of insulation and the method of application of the test are controlled to be the same for all such tests.
5. It is not possible reliably to "predict" breakdown of insulation under such tests, from leakage-current characteristics.
6. Distribution of electric stresses under direct-voltage tests are different from those under alternating-voltage stresses. The results of direct-voltage tests, therefore, cannot reliably indicate whether an insulation has sufficient voltage strength for operating under alternating voltage stresses.

The authors' evaluation, that separates the reliable aspects of direct-voltage tests from limitations of such tests, based on results of tests on complete machine windings, complements a similar evaluation of these tests, based on results of tests on special sheet insulation samples (see reference 11 of the paper).

Measurement of Ionization During Direct-Voltage Tests

The additional measurement of ionization, during direct-voltage tests, as recommended by the authors, seems to increase the reliability of interpretation of such test data. It also reduces the possibility of damage to insulation under test, for reasons briefly discussed in the following and described in detail in reference 11 of the paper.

Under direct-voltage stresses, on special sheet insulation samples tested between circular brass electrodes, extensive regions of the samples had binder material degraded by "the current density-thermal-ionic mechanism." Such a degradation of the binder was caused by interlayer ionization activity inside the samples. Under alternating-voltage stresses, no such regions of degraded binder were observed. Surface ionization was very profuse and evenly distributed, under alternating-voltage stresses, while it was almost totally absent under direct-voltage stresses. Such differences in the nature of ionization, under the two types of stresses, resulted in degradation of insulation, under direct-voltage stresses, in regions where alternating voltage had little immediate effect.

The test data by Odok and Soelaiman indicate that some ionization preceded direct-voltage breakdown of insulation. While this does not conclusively prove that the insulation was being progressively degraded by the ionization activity, it does indicate that by stopping direct-voltage tests, when ionization is initiated, the possibility of damage to insulation, under such tests, may be reduced or eliminated.

Summary

1. The authors' recommendation of measuring ionization along with leakage current and polarization index, during direct-voltage tests, will increase the reliability of interpretation of such data, and it will reduce the possibility of damage to insulation under test.

2. Their evaluation separates the reliable aspects of direct-voltage tests from limitations of such tests.

This paper is a very timely and significant contribution to the field of insulation testing.

A. W. W. Cameron (Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada): The authors should be congratulated on an interesting study, carefully reported. High-frequency discharge detection, as they have developed it, shows promise of being a useful adjunct to d-c high-potential testing of some insulations. It is hoped that further study and experience will show the areas in which it may most effectively be applied.

One question I would like to ask is whether the authors have encountered, in their experiments, pulses, resembling ionization pulses, which are associated with the absorption current, and also with the corresponding discharge current. Apparently some of the polarization processes proceed in discrete bursts. In 1950 I observed these pulses by a circuit similar to Fig. 4, and saw that they recurred very rapidly after a stepwise increase in direct voltage, becoming progressively less frequent as the absorption component of current decayed. Upon removing test supply, and short-circuiting the specimen, pulses of similar amplitude but opposite polarity accompanied the discharge current in the same manner, diminishing in repetition rate as the d-c microampere reading decreased. I was able to puzzle my colleagues by showing them what seemed to be ionization pulses in insulation at zero voltage, although such a process should not, of course, be called ionization. This phenomenon has also been reported by Dr. Wichmann (see reference 1 of his discussion).

It would seem that some care would be required in discriminating between these absorption pulses and the discharges which the authors wish to utilize.

Fig. 15(A) shows discharges from 5 kv upwards, suggesting some difficulty in judging when intensity is dangerous. On the other hand, a sixfold increase in direct current between 7 kv and 9 kv looks like a rather clear indication. Fig. 15(B) shows that external interference may be a complication in assessing discharges; again, more than five times increase in current or four times decrease in resistance between 1.5 kv and 2.0 kv is hardly to be overlooked. I think these current indications call for reconsideration of the statement that leakage current gives "practically no information."

The authors' findings refer to d-c machines, in which leakage is complicated, or even dominated, by commutators and other components offering surface leakage paths. The insulations investigated seem also to have had very little absorption compared with insulations usual in high-voltage stator windings. Therefore, I am wondering to what extent these findings will be applicable to the higher-voltage machines.

Regarding Appendix II on absorption, I believe that the major absorption in high-voltage machine insulations, which may take hours to diminish to insignificance, has usually been ascribed to interfacial polarizations.

The statement that "high-voltage d-c testing is inherently better suited for d-c machines and equipment than for a-c machines" is one which I question. The armature coil of an operating d-c machine is subject to alternating voltage superimposed on a direct voltage. It is insulated generally in the same manner as an a-c machine coil. It is my belief that a testing voltage should be selected according only to the condition which it is required to detect, and that our commonly assumed relations between testing and operating voltages are liable to be illusory. I heartily echo the authors' warning against the use of magic single values of "d-c-a-c ratio."

Direct-voltage tests have been found suitable for detecting cracks, fissures, delaminated areas, and other conditions involving paths through high-voltage machine insulations. This paper indicates that d-c tests also detect adverse conditions in open creepage paths. I am rather surprised that some failures of traction machines are not ascribed to fissuring of insulation, and to tracking on creepage paths. Also, I would not have expected many faults in these lower-voltage machines to be caused by ionization. In my experience, d-c tests have not indicated thermal degradation, unless it was so far advanced that ionization or cracking presented a through path.

The fears which the authors express regarding d-c leakage-versus-voltage tests have not been justified in my 7 years' experience of their application to high-voltage a-c machines. These tests have been reliably nondestructive. If the authors' discharge detection method can confer any additional significance upon such tests, may I be the first to welcome it. Figs. 1, 2, 3, and 6 indicate experience consistent with my own. I cannot agree that a d-c test on "good dry insulation" involves any danger of breakdown, because if the insulation is good and intact it will very easily withstand a suitable maximum test voltage. If the insulation is wet, however, this should be detected by a simple megger test before a high-voltage test is attempted.

Appendix III appeals to me as giving a very good picture of a mechanism of d-c conduction across an insulating surface with increasing voltage. The point which I think has been missed here is that similar mechanisms would be expected to apply also to conduction along the surfaces of fissures through the insulation. Therefore, this may well be an explanation for high-voltage d-c tests detecting such fissures reliably and nondestructively.

The effect of moisture is obviously important in such surface and fissure conduction mechanisms. Fig. 12 does not specify humidity conditions, but the peculiar leakage-current behavior at the higher temperature may be due to drying up, by the current, of small amounts of moisture. In my experience, consistent and significant d-c nondestructive tests are obtained when windings are cooled to ambient temperature in air of normal humidities. Adverse reports seem to have been based upon tests made on warm or very dry windings.[1]

Regarding applications to a-c testing, I would say that my colleagues are using a discharge detector to determine ionization inception and extinction voltages in large rotating machines. A few minutes' running on open circuit, with varying excitation, is all that is required. This seems to give a good indication of void formation, and of the general loosening of the insulation structure which gives a predisposition to turn-to-turn faults.

Attention is invited to a paper by F. S. Oliver which contains interesting information on discharge measurements.[2]

In conclusion, I am interested in the possible benefits of quantitative analysis of discharges, and I hope the authors will be encouraged to continue these studies and to favor us with further reports. I thank them again for a wealth of valuable data and a thoughtful paper.

REFERENCES

1. EXPERIENCE AND DEVELOPMENT IN NON-DESTRUCTIVE D-C TESTING FOR MAINTENANCE OF HIGH-VOLTAGE STATORS, A. W. W. Cameron, A. M. Sinclair. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 75, Apr. 1956, pp. 201–10.

2. MEDIUM-VOLTAGE A-C TESTING OF ROTATING MACHINERY INSULATION, F. S. Oliver. *AIEE CP 58-203* (available on request).

E. B. Curdts (James G. Biddle Company, Philadelphia, Pa.): I agree with the authors of this paper that tests on the insulation systems of low- and medium-voltage equipment which includes the detection of corona or gaseous ionization should be useful in evaluating the condition of such systems. Whether a supplementary ionization test can be economically justified as a maintenance procedure is open to question and should be further investigated. The skill required of repair-shop personnel in interpreting test results may present a problem of economics, and electrical noise causing unpredictable distortions of the cathode-ray oscilloscope pattern may point to the need of effective input filtering and specimen shielding which may add to this economical problem.

I believe that something can be learned by the ionization tests suggested by the authors and further investigation is certainly justified. However, there are several categorically made statements in this paper which should be questioned.

One of these concerns the megohmmeter test described in the opening section of the paper. The insulation-resistance test as made by such instruments is not necessarily a low-voltage test for equipment considered in this paper. Numerous motor repair shops and others use megohmmeters rated up to 5,000 volts with provisions for selecting several voltages below the maximum.

Second, regardless of statements made by the authors, insulation-resistance tests have been for years and are continuing to be used with a great deal of success in preventive maintenance activities. Several methods are in use, their choice depending somewhat on economics. The simplest of these, the 1-minute test at a voltage usually low compared to the rating of the specimen, very faithfully reveals the presence of dirt, moisture, and carbonized paths. Its usefulness is limited only because it cannot distinguish between relatively harmless generally distributed leakage conditions and localized incipient faults. Insulation-resistance tests made with several voltage steps ranging from below to somewhat above the rated voltage of the equipment under test can give more information than a single-voltage

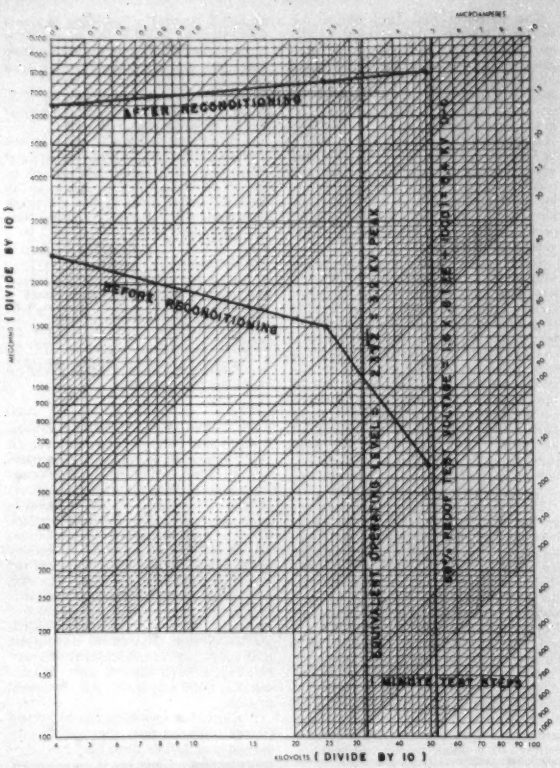


Fig. 19. A 350-horsepower 2,300-volt 3-phase motor armature before and after reconditioning

resistance measurement concerning the localized characteristics of abnormal leakage current.

The authors give an example of the unreliability of low insulation-resistance values on diesel-electric locomotive equipment resulting from tests made in very humid atmospheres. If such humid atmospheres are not representative of operating conditions because the insulation is operated at temperatures in excess of the dew point, then the test is not realistic in the first place. On the other hand, if the same conditions of humidity prevail in service as exist during tests, then the measurement tells us from experience what values can be tolerated in service when caused only by generally distributed leakages. However, for maximum reliability higher step-voltage insulation-resistance tests can also be made in an effort to reveal any incipient faults that may exist. I doubt very much if ionization tests, under the high-humidity conditions cited, can do more. In fact the authors make the following statement in their paper: "Discharges start in a moist, contaminated insulation system at a much lower voltage level than the actual breakdown voltage. The test will therefore usually be stopped long before a breakdown can be anticipated. This may result in overcautiousness so that the anticipation of breakdown will be too far on the safe side. This circumstance, which can to a great extent be corrected by experience and skill in testing, is no drawback because moist and contaminated insulations will usually be treated in some way to remove polar materials before being put into service and subjected to normal operating stresses."

I am afraid that the authors, in their understandable enthusiasm for the ionization test, have become unduly critical of the voltage-current or voltage-resistance method of test. They say that it alone cannot tell anything about the presence of moisture or contamination, and supply practically no information about the insulation condition. Our experience definitely does not bear this out. We have found that excessive moisture and dirt, or what might be called "poor housekeeping," can be easily detected by the voltage-resistance or voltage-current method. Fig. 19 is a typical example of conditions as found and the results of reconditioning a 2,300-volt motor armature. I can find nothing in the data given in this paper to indicate that the ionization test would be a more convincing detector of dirt and moisture than the voltage-resistance or voltage-current test. The nonlinear behavior of the currents given in Fig. 15 of the paper certainly seem to be as informative as the corona patterns. Incidentally, the nonlinearity of the current given in these figures can be easily appreciated without resorting to curves with ill-chosen co-ordinates which is one of the objections raised by the authors to the analysis of leakage current. Furthermore, it is not necessary, in the analysis of leakage-current behavior, to apply each voltage step long enough to establish a "steady-state current." The rising characteristics of the insulation resistance of the curve shown in Fig. 19 after the armature was reconditioned is caused by the remanent, superimposed absorption currents at the end of each 1-minute step. Understanding the reason for this rising behavior makes it unnecessary to eliminate it by increasing the time for each voltage step.

The authors state that repeated voltage-resistance or voltage-current tests are not usually reproducible. We have found the contrary to be usually true, particularly in the case of clean dry windings. Very dirty or wet windings will sometimes change their leakage-current behavior from one test to another. This being so, and in view of the authors statements to the effect that the corona inception voltage may be very close to the actual breakdown value for clean dry windings, I do not agree that it is necessarily true that the corona test is less destructive. I do not believe that either test is destructive if conducted properly, even on insulation systems involving localized areas which are in a marginal condition.

In connection with the authors' statements concerning absorption tests, I would like to make a few brief comments. First, the absorption or time-resistance test as usually made in maintenance practices is not done to determine the absorption characteristics of the insulation, but rather to determine the relative magnitude of the leakage current. In other words, the polarization index provides an indication of the relative dryness and cleanliness of the insulation, again a measure of "good housekeeping." If the test voltage is raised to levels that cause leakage-current nonlinearities then the ratio of absorption to leakage current changes which in turn changes the polarization index. The perfectly flat curve for wet insulation given in Fig. 10 to me seems odd indeed, and I question its validity.

In closing this discussion I would like to raise an academic question concerning the authors' repeated reference to "polar materials such as dirt and moisture," and that "the presence of polar molecules causes the insulation to behave more like a pure ohmic resistance." I do not question the fact that the high dielectric constant of pure water is largely due to dipole polarization, but I do not see the point of dwelling on this since the high conductivity of contaminated

water, which we would expect to find in insulation that has seen service, may outweigh its significance. I have found no evidence of a predominate reversable energy in dirty and wet insulation after having been subjected to a direct voltage. I would appreciate the authors' further comments on this phase of the subject.

The authors also make several contradictory references to interfacial polarization which leaves me somewhat confused. In one place they attribute the capacitance of a machine winding wholely to interfacial polarization when measured at power frequencies, and in another they say the interfacial polarization contribution to the d-c dielectric constant is complete in less than a minute. In a third place they say that interfacial polarization plays only an insignificant part in current behavior from the standpoint of the time constant with which the steady-state current is established. Unless I have misunderstood the meaning of these statements I must disagree with all of them. For example, the 60-cycle capacitance of the usual machine winding is made up largely of contributions from the so-called infinite-frequency polarizations as well as dipole and possibly some interfacial absorption. Since the time constant of the interfacial type of absorption is often found to be longer than 1/120 second its contribution to the 60-cycle capacitance may not necessarily be significant. The infinite-frequency polarizations, such as the electronic and atomic types, have very short time constants in themselves and, unless the resistance of the test circuit is quite high, should be complete within a minute. It is to avoid the measurement of these polarizations that one minute has been chosen as the first time interval for polarization index measurements. The interfacial absorption can be observed for hours.

N. Klein (Israel Institute of Technology, Haifa, Israel): This paper presents observations and testing proposals of basic importance and it is of great interest to ascertain their range of validity at higher voltages and for other equipment and materials. It is a merit of this paper that the authors clearly differentiate between the information obtainable from the leakage current, absorption current, and discharge inception tests respectively.

The essential point appears to be the recommendation to introduce discharge inception measurements on d-c tests. This kind of test is customary with a-c apparatus. In the case of alternating current however, observations show completely different behavior, as frequent discharges start at comparatively low voltages, when breakdown follows only after a long time, or not at all. On the other hand, in the case of direct current the authors' observations indicate that frequent discharges start at much higher voltages than with alternating current and with dry insulations breakdown follows soon.

It will be important to investigate the reasons for these differences in behavior and some comments will be made here in this connection. The breakdown tracks observed by the authors developed along external surfaces and across the insulation. Let us consider the part across the insulation and the consequences of discharges in cavities occurring in the insulation. According to a simple model, it is often assumed that when the voltage is increased to V_i and a discharge occurs in the cavity, positive and negative charges become attached to opposite surfaces of the cavity and the electric field vanishes in the cavity owing to these charges. On increasing the voltage to $2V_i$ discharges occur again, the charges are added at the cavity surfaces to the charges of the previous discharge, and the field in the cavity vanishes again. This process is repeated on further voltage increase and it is expected that infrequent discharges will be observed on the oscilloscope at about equal time intervals.

The authors observed however a sudden increase in the repetition rate of the discharges at a voltage which is a multiple of V_i. In the light of the model of discharge sequence described it would appear then that the charges attached to opposite cavity surfaces diminished and it could be assumed that they were neutralized, or conducted away. In the case of damp insulations the first assumption seems likely. In the case of dry insulations, however, neutralization of negative charges does not seem to be plausible and the question arises: "How are the observations connected with an increase in the conductivity of the insulation?" Consider the electrons of the discharge. These impinge upon the cavity surface with energies up to 100 electron volts and to a great part penetrate into the solid, where they ultimately become attached to trapping sites. On repeated discharges trapping sites increasingly closer to the electrodes become filled. When with increased voltage the discharge frequency suddenly increases, it is suggested that several processes can take place in the solid. Possibly the trapping centers are filled to such an extent that electrons from further discharges move to the electrodes without being trapped. Possibly electrons are liberated owing to the high electric field by various mechanisms, e.g., from traps. Processes connected with positive ions at the cavity surface are even less clear and neutralization by electrons from the solid might be instrumental. In any case a sudden increase in the conductivity of the solid adjacent to the cavity seems to be connected with the increase in the frequency of discharges and consequent breakdown and experiments will have to show whether the afore-mentioned or other processes cause the conductivity increase.

Such a mechanism of d-c breakdown appears to be different from those known and also to a certain extent from the mechanisms of a-c breakdowns, caused by discharges in cavities. It would be important to ascertain to what extent the development of breakdown on external surfaces is influenced by trapping centers. It seems from these observations, and mainly from such with alternating current, that the trapping sites, which constitute a very small fraction of the insulating materials, can influence the breakdown strength. The question arises then as to whether a control of the trapping centers could not help in the improvement of insulating materials.

The authors consider that the use of discharge detection for a-c conditions on alternating current is more involved and it is not clear what value information obtained on d-c discharge inception tests can have for estimating insulation quality on a-c operation.

It is stated in the paper that practically no information can be obtained about the insulation condition by the leakage-current characteristic in itself. This is most probably true with regard to the imminence of breakdown, for which discharge inception and absorption current tests, as recommended by the authors, can give the information. On the other hand, leakage-current tests instruct about the conducting properties; e.g., the leakage current versus temperature measurements recorded in Fig. 11. When plotting the logarithm of the leakage current against the reciprocal absolute temperature, it is interesting to note that for various voltages parallel straight lines can be drawn through the experimental points with quite good accuracy.

It would appear that the authors have made an important contribution to the subject of testing insulating materials and are to be congratulated. It is hoped that this work will be continued experimentally on apparatus and on samples of insulating materials under carefully controlled conditions.

E. W. Bultman and **C. W. Martin** (Baltimore and Ohio Railroad Company, Baltimore, Md.): We feel that this paper is very interesting and informative. It is another link in the continuing investigations that are being made in d-c testing of insulation systems in low-voltage d-c equipment.

We are engaged in the maintenance and repair of diesel-electric locomotive equipment. For several years we have been conducting an intensive investigation from all sources available to us on the subject of insulation testing and the evaluation of these tests. Present uses of d-c testing, while considerably improved over former methods of testing insulation, still are not fully informative.

The concept presented raises the hopes for improving test programs on the railroads, but it also raises some questions. The data presented in this and previous papers are always concerned with the insulation resistance levels that we as operators never see or attempt to achieve as they are not economically feasible. In our own motor shops equipment is released for service with several microamperes leakage current and the equipment operates successfully. Also, we have received, repaired, and rewound traction motors and main generators from leading manufacturers which, when received, meggered 5 megohms with a 500-volt megohmmeter yet they carry a new motor warranty. If we used an ionization-detection method to test this equipment would we be able to release it for service?

What will be the ionization pattern for a traction motor's insulation system which is very dry and clean but which has a puncture? This type of fault on a d-c test usually has a gradual smooth straight curve to a given impressed voltage and then breakdown comes without warning. No knee occurs in the curve for this type of failure.

Investigations and discussions with other railroad personnel indicate that locomotives are being operated with power circuits having insulation resistance in the order of 5

milliamperes at 1,500 volts d-c. Have any tests been made at this level of insulation resistance? If so, what were the results?

What voltage test level and leakage-current values do the authors recommend for testing complete locomotive power circuits, individual traction motors, and main generators on the locomotive, and the same equipment in a rewind shop?

It seems the most practical use of the ionization theory is in predicting the breakdown voltage and stopping the test prior to reaching this point. Do the authors agree?

We would appreciate a discussion of the availability and costs of the component parts of the test equipment used by the authors and if such a test set is or might become available commercially.

G. L. Moses and **J. S. Johnson** (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The authors should be congratulated on the preparation of a very thorough and well-documented paper on the subject of d-c high-potential testing of insulation systems for low- and medium-voltage d-c equipment. The authors make a very strong case for the use of ionization measurements to determine the condition of the insulation with respect to dirt and moisture contamination, and the probable level of insulation strength on clean and dry insulation, for the types of insulation systems which they have investigated. In connection with the development of applicable test methods for insulation evaluation, caution should be exercised in assuming general applicability of a test method which has shown a great deal of promise when used for limited types of equipment. While we have seen some evidence that ionization phenomena precedes breakdown on higher voltage equipment, the correlation has not been as marked as the authors have experienced in connection with the insulation systems which they have studied. This is probably due to the difference in the insulation system, the thickness of insulation used, and probably to a large extent, the conditions of use. The components of railroad equipment are subject to much more adverse treatment from the standpoint of exposure to contaminating influences than are the large turbine generators, which operate in a clean and dry atmosphere. It therefore must not be assumed that the same measure of success would be experienced in making similar tests on this type of equipment.

It is probably true, however, that more emphasis should be placed on the detection of ionization pulses and their relative frequency with increasing test voltages, even for this type of equipment. Studies along these lines were reported as early as 1941 when the Edison Electric Institute Subject Committee on field testing the generator insulation reported results of destructive tests on 106 large generators which were about to be rewound. Among these a series of ionization studies were reported for eight generators, with the conclusion that the ionization measurements made did not appear to be effective in indicating the existence of isolated points of weakness in the windings tested. There is certainly room for further investigation, however, of this general field.

The authors very correctly point out that the shape of leakage-voltage characteristics may be greatly distorted by the arbitrary selection of scales used in plotting these characteristics. We have found that in interpreting leakage-voltage characteristics of machines, the basis of comparison should be the leakages which are normally expected for machines of the type being tested. If the absolute magnitude of the leakages or the rate of increase are high, when compared to those which would normally be expected, this is investigated and the possible causes determined. Also, care is exercised in the selection of scales so that unusual leakages will be readily indicated. Here, again, it should be pointed out that in the testing of high-voltage insulations, moisture contamination is relatively unimportant as compared with the equipment which the authors have studied.

While there may be some question as to the general applicability of this method the authors have made a very constructive contribution to the problem of anticipating electric failure at least for the equipment which they have studied.

Direct-voltage testing has been used quite successfully for some time as an over-voltage proof test to demonstrate the adequacy of a winding to withstand a certain direct-voltage level. There have been many attempts made to use leakage current versus direct voltage to provide a forecast of impending failure but with somewhat mixed success. The authors' proposal to observe the ionization discharges while applying a direct-voltage test certainly warrants further study and trial.

The method proposed by the authors should be employed by several investigators on a variety of electric equipment with different kinds of insulation. It is recommended that the authors make statistical studies on a large number of laboratory-prepared specimens and report their data at a subsequent meeting of the Institute. If this method proves to be as useful as the authors' preliminary study indicates, it would truly be a significant contribution to the art of testing.

A. M. Odok and **T. M. Soelaiman**: We wish to thank all who contributed discussions for the very gratifying interest shown in this paper. So wide a response, it seems, is an indication of dissatisfaction with the present status of insulation-evaluation methods and of open-mindedness toward new approaches in this field.

Before considering the discussion in detail, it seems appropriate to stress some facts concerning the general aspects of the method described in this paper:

1. Contrary to first impression, the use of this method involves no more skill than does the step-voltage d-c test. In fact, it has been found unnecessary to plot the leakage-current characteristic during the test since a good picture of the insulation condition can be obtained without checking the slope of this curve. The oscilloscope need not be an elaborate universal type, but can be simplified as single-purpose instrument, thus reducing weight, size, and complexity.
2. Since the noise current is picked up by a series resistor or choke no tuning circuit is used parallel to the test sample. Also, the method proposed in this paper differs from ionization-detection methods using probes.
3. Field-testing experience on a prototype equipment with built-in oscilloscope indicates that radio noise causes no particular difficulty with this type of noise-current pickup, except under highly adverse conditions, such as closeness of arcing devices, etc.

In response to Dr. Wichman's very thorough discussion, it may be pointed out that disturbances by radiated or line-transmitted noise have not been a particularly difficult problem in these investigations. If noise is displayed on a Lissajous figure where the horizontal deflection is, for example, the ripple voltage of the rectified direct voltage, the line pickup and interference are avoided, since such noise then shows up in a radial direction on the Lissajous ellipse, whereas the discharge noise appears as vertical pulses on the screen.

The insulation system tested was class *H*. However, we would like to point out that we have not tested any high-voltage apparatus such as that to which Dr. Wichman refers.

Mr. Simons' experiences with high-voltage d-c tests seem partially to support the working hypothesis that d-c breakdown always is preceded by noise current. It would be interesting to know what ionization-detection method was used in the tests he mentions. In the absence of this information, it may be assumed that probes external to the test circuit have been used, so that only local discharges were picked up.

It is interesting that Mr. Simons reports success with a-c ionization detection similar to that reported in this paper for d-c tests.

Mr. Bhimani, in his discussion, attempts to itemize the conclusions reached in this paper. We would like to thank him for this contribution to a further clarification of our stand, but feel obliged to comment on some points which might be misunderstood:

The first two "limitations" which he lists seem to imply that some other single test exists which will supply all this information conclusively. Since the purpose of the test is to detect weaknesses of the insulation to electrical stresses, the method need not be sensitive to every kind of abnormity in the system unless it represents an inherent dielectric weakness.

Although the temperature is significant for the absolute level of leakage current, it does not affect the application of the described method, as long as it is constant and within reasonable limits.

Comment should also be made on item 6 under "Limitations" where the difference in distribution of stresses under direct and alternating voltages is brought out. While we agree with the principle stated therein, we wish to point out that no attempt was made to correlate or compare the value and effects of the two types of stresses, and the two types of tests.

Mr. Cameron's question regarding ionization pulses occurring during discharging of the insulation system seems to have found a plausible explanation in Dr. Wichman's discussion. We have observed this phenomenon without attempting to use it to differentiate between void discharges and other types of discharges as reported by Dr. Wichman.

Mr. Cameron and Mr. Curdts point to

the increase of leakage current in Fig. 15 and ask whether this is not as good a measure of approaching breakdown as the noise current. Unfortunately these two examples are not typical of all possible conditions of the insulation. Experience indicates that there is some connection between increasing leakage current and intensive noise current. The curve in Fig. 17 of Dr. Wichman's discussion also points out the parallel between the dissipation factor and the frequency of d-c pulses. Combining this with, on one hand, Dr. Klein's attempt to explain the nature of pulses, and, on the other hand, the experience of Mr. Cameron and the authors with pulses during charging and discharging, one is led to speculate that the space-charge ionization takes place in discrete bursts. In other words, the noise observed consists of Barkhausen pulses similar to the domain orientation phenomena in magnetics.

Mr. Cameron questions the statement that "high-voltage d-c testing is inherently better suited for d-c machines and equipment than for a-c machines." Despite the fact that d-c machine armatures are subjected to a combined stress composed of an alternating voltage superimposed on a direct voltage, we think the statement holds for the following reasons:

1. There are parts in a d-c machine which are exposed only to direct-voltage stresses.

2. No information is yet available as to the effect of a pulsating unipolar voltage on insulation. However, this type of stress and a constant-voltage stress have at least one point in common: the polarity does not change, therefore the ionization mechanism is different from that under alternating-voltage stresses.

Mr. Bultman and Mr. Martin report experiences with resistance testing which are in perfect conformity with those of the authors. The answer to the question as to whether the proposed method would help toward a more definite screening of traction motors than is possible by resistance measurement is indicated by our field experience. Even though the proposed method is still being evaluated in service shop use, experience thus far shows that the additional information obtained is of considerable help in determining the cause of low resistance and deciding whether or not it represented a weakness of permanent nature which would lead to breakdown under service conditions. Of course, more field data are needed to answer the specific questions these discussers have asked. Generally speaking, neither a low nor a very high resistance, as measured with a low voltage, is useful unless its cause is known. Undoubtedly, both leakage- and absorption-current measurements supply useful information. Our experience indicates that the proposed test method is quicker and supplies as much or more information than leakage-current and absorption tests.

Comments on some other questions asked by Mr. Bultman and Mr. Martin are:

1. A puncture will be detected if it is at a point so highly stressed as to give rise to ionization of the surrounding air or if its surface is sufficiently contaminated to cause surface tracking.

2. Resistances as low as the 300,000 ohms mentioned have been encountered. Most often this occurred with moist insulation and the diagnosis of humidity usually helped to classify the insulation as not permanently weakened.

3. The authors are not in a position to recommend general rules as to voltage and current levels in traction equipment, since these depend on nominal voltage, size of equipment and many other factors. These values are not highly significant in nondestructive testing since the real voltage limit is found by observing the conditions which will start damaging the insulation.

4. Certainly the most practical use of the noise-current detection is to give a more dependable prediction of breakdown than could be obtained with other methods. Other useful features might, however, prove even more important in service shop applications as described in the paper; these concern mainly the condition of the insulation system.

We are grateful to Mr. Curdts for having called our attention to various inconsistencies which have since been corrected in the text.

Dr. Klein's attempt to explain the mechanism of discharges observed under constant direct voltages has much appeal and we agree with the necessity of carefully controlled experiments to check his hypothesis. We also hope that other workers will gain more experience with the method described in order to prove the extent to which it will be of practical value to the field engineer.

We wholeheartedly agree with Mr. Moses and Mr. Johnson that too early a generalization of the presented method to high-voltage equipment should be avoided. This caution was implied in the title of the paper and the specific reference to the types of machines tested. It is to be hoped that other investigators will supply information regarding the usefulness of this method in other types of equipment.

The leakage-current characteristics would be more informative and easier to interpret if normal leakage-current curves were available for each piece of equipment to be tested. While this might be economically justifiable for large machines, it is doubtful whether a user of a large number of small machines would be willing to keep accurate records of their individual insulation resistance. Furthermore, tests would have to be made under more carefully controlled conditions than those usually associated with periodic inspection testing of such machines. It should here be pointed out that, as reported by Mr. Bultman and Mr. Martin, a change in leakage resistance has not been found to be a good measure of the operability of traction motors.

A Recommended Program for Resistance-Welding Instrumentation

A. DIXON
MEMBER AIEE

In a study of the characteristics of known basic variables in resistance welding, a preliminary conclusion indicated that the magnitude of variation in each function in the order of the severity of their variance is: weld current, most variable; weld force, less variable; and weld time, least variable. Consideration of these conditions indicated the desirability of a continuous system for monitoring of current and force, with periodic checking of the weld-time function.

Upon evaluation of existing methods, the following mediums were considered of acceptable reliability for instrumenta-

THE PROBLEM of developing a practical method of instrumentation for production use at the Westinghouse Electric Corporation has been advanced by the combined interest and efforts of research, development, and manufacturing people within the company. The use of instrumentation for resistance welding involves wide economic considerations in product quality and production costs and has thereby greatly stimulated this interest.

In the process of this work many mediums of instrumentation were compared and evaluated in the manufacturing laboratory and production departments. The standard for comparison was the very accurate magnetic oscillograph. The outcome of this effort was the adaptation of instrumentation methods having simplicity as a keynote, to enable the effective use of such procedures at the resistance-welding operator level and at a minimum cost per unit machine.

Paper 58-210, recommended by the AIEE Electric Welding Committee, and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 16, 1957; made available for printing January 7, 1958.

A. DIXON is with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

tion of variables on single-phase equipment:

1. *Weld current.* A back-stop current ammeter and a primary current transformer for weld-current amplitude setting of machine and monitoring. This is supplemented by the brush amplifier and oscillograph where modulated current waveforms are used (current rise and taper characteristic). Relative current amplitudes are thus observed and recorded. Certain Military Standard resistance-welding work includes this as an inspection and quality control requirement.

2. *Weld force.* A pressure gage and pneumatic pressure regulator, which is standard equipment on most resistance-welding machines, is used for setup and monitoring. Where schedules require the use of dual force systems, a strain gage and amplifier are used in conjunction with the brush oscillograph. The specific purpose in this case is to determine correct timing for the forge-pressure delay function.

3. *Weld time.* This variable is predetermined by an initial calibration test, and is recorded on standard calibration and schedule charts. Periodic maintenance checks are the maximum requirement. However, when using brush instrumentation for current measurement, a monitor and permanent record are also provided for the weld-time variable.

In considering the over-all problem, other factors such as condensed welding schedules and recommended resistance-welding electrode combinations were determined. Maintenance programs have also improved considerably with established calibration procedures and recorded machine calibration charts of instrumented variables.

The simplicity of the program as outlined in the following has established economic advantages, and presents a practical manufacturing engineering approach to instrumentation of and improved resistance-welding operations at Westinghouse.

Background

Resistance-welding operations within the company are in many diverse forms, which impose varying degrees of instrumentation requirements, many of which have been standardized and recommended welding schedules applied. These methods, which will be described in the following section, may be categorically classified as follows:

1. Machine preparation, or equipping the resistance-welding machine for instrumentation.
2. Making machine calibration records.
3. Preparing welding schedule charts.
4. Resistance-welding instrumentation for aluminum work.
5. Miscellaneous applications.
6. Effective use and application of instrumentation.

A preliminary survey of the basic variables which can be readily instrumented indicates a maximum need for the instrumentation of weld current, since this quantity is readily affected by factors which are not within the scope of simple control. Such factors include line-voltage changes, the addition of weld material in the throat of the machine, and physical changes on the secondary electric circuit of the welder transformer due to operator changes in welding setup of electrodes or fixturing.

Although the variable of weld force does not have the severity of change characteristic of weld current, the facility for continuous monitoring of this variable is provided as standard equipment on most resistance-welder machines. This consists of a pneumatic-pressure regulator and a pressure gage which indicates the air pressure to the air cylinder of the machine. An occasional reference to actual air-gage readings by the operator and the posted calibration or schedule chart, detects any change due to inconsistencies in pneumatic systems and supplies.

The weld-time variable is the most consistent factor and therefore, for most work, does not require the continuous use of a monitor-type instrument by the operator. An exception to this is where an oscillographic record is required in the certification of work schedules as an inspection or quality-control requirement. This is frequently the case where a modulated current wave is used in the welding schedule.

A system of providing machine calibration records and charts offers the following advantages:

1. It enables better scheduling of work within welder machine capacities, thus obtaining good consistency machine by machine.
2. It enables the use of a process specification, which includes the recommended welding schedules and can be specified on engineering design drawings.
3. It reduces machine setup time and places the initial responsibility for correct welding procedures within visual supervisory control.
4. It encourages standardization and the use of recommended practices in resistance-welding production operations, reflecting considerable savings in labor and material and reduced waste in defective work.
5. Machine calibration record charts enable the detection of failure and the rapid localizing of machine defects. A knowledge of normal machine performance in the adjustable variables of time, force, and current offers a convenient standard for

Fig. 1 (left). Current transformer

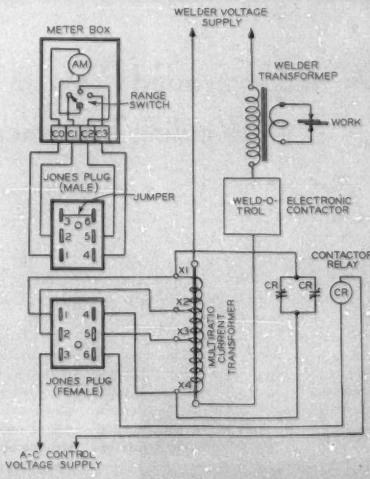


Fig. 2 (right). Weld current meter diagram of connections with automatic current transformer short-circuiting relay

Weld current (amperes) = ammeter reading × welder transformer ratio × current transformer ratio

comparison at the time of failure or suspected failure.

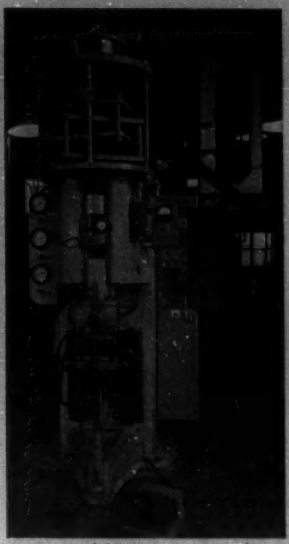
Calibration and Welding Procedures

Machine Preparation

The first step in preparing a specific resistance-welding machine for instrumentation is in the selection and installation of the primary current transformer. Two standard styles of current transformers provide the required range of primary current coverage. The first style has ratios of 200, 400, and 600 to 5, and the second style has ratios of 1,200, 600, and 300 to 5. A typical current transformer for this purpose is shown in Fig. 1. A permanent installation is preferred because of convenience and low cost. To determine the correct size of current transformer required, a reasonable approximation of the maximum value of primary load amperes can be made from the nameplate output rating of the resistance-welding machine in secondary amperes divided by the welder transformer ratio.

The second step is in the selection and mounting of a back-stop current ammeter. When possible, a permanent installation is also desirable. However, if a portable meter arrangement is desired, a "Jones"-type disconnect plug can be used, as illustrated in Fig. 2. This system will enable the portable use of one ammeter on N number of machines. The system incorporates an automatic short-circuiting feature for the current transformer secondary, which is done by an electric contactor relay when the plug is disengaged. The most widely used back-stop current ammeter, shown in Fig. 3,

Fig. 3. Weld current meter



Federal Machine and Welder Company

Fig. 4. Spot-projection welder with correct meter installation

Table I. Welder and Meter Data for Precision Rocker Arm 250-Kva Spot Welder

Welder Transformer Data			Weld Current Meter Data	
Primary Tap	Secondary Volts	Welder Transformer Ratio	Range Switch Position*	Meter Multiplier
Lo-1	7.3	60.25 to 1	1	0.5
Lo-2	7.9	56.7 to 1	(Use up to 300 amperes primary current)	
Lo-3	8.5	51.8 to 1		
Lo-4	9.2	47.8 to 1	2	1
Hi-1	10.5	41.9 to 1	(Use up to 600 amperes primary current)	
Hi-2	11.6	37.9 to 1		
Hi-3	13.0	33.84 to 1	3	2
Hi-4	14.7	29.9 to 1	(Use up to 1,200 amperes primary current)	

Primary volts = 440

$$\text{W.T. Ratio} = \frac{\text{Primary Volts}}{\text{Secondary Volts}}$$

Example: Ratio of tap Lo-4 $= \frac{440}{9.2} = 47.8$

To set current meter for required secondary current:

$$\text{Meter reading} = \frac{\text{Required secondary amperes}}{\text{Meter multiplier} \times \text{welder transformer ratio}}$$

Example: To set weld current meter for 26,100 amperes

$\frac{26,100}{29.9\dagger} = 874$ amperes, primary current

Use range switch on no. 3 position

$$\text{Meter reading} = \frac{26,100}{(2 \times 29.9\dagger)} = 437 \text{ amperes}$$

Set pointer stop for 437

* Determine primary current by dividing secondary current by the welder transformer ratio for primary tap used.

† Hi-4 primary tap setting on welder.

Table II. Short-Circuit Weld Current Calibration for Precision 250-Kva Rocker-Arm Spot Welder

Primary Tap	% Current Dial	Primary Amperes	Secondary Amperes
Hi-4	20	264	7,900
	30	372	11,130
	40	540	16,150
	50	684	20,450
	60	888	26,600
	70	1,057	31,600
	80	1,249	37,350
	90	1,512	45,200
	100	1,583	47,300
Hi-3	20	216	7,330
	50	540	18,300
	100	1,249	42,250
Hi-2	20	164	6,370
	50	444	16,820
	100	1,080	35,100
Hi-1	20	132	5,530
	50	360	15,100
	100	888	37,250
Lo-4	20	108	5,170
	50	276	13,200
	100	660	31,550
Lo-3	20	96	4,975
	50	252	13,070
	100	570	29,550
Lo-2	20	84	4,760
	50	210	11,900
	100	516	29,250
Lo-1	20	72	4,340
	50	180	10,850
	100	438	26,400

Calibrated with 1¹/₂-inch-diameter copper bar.
Throat dimensions: opening, 8 inches; depth, 24 inches.
No-load volts, 446.
Full-load volts, 440.

Fig. 5. Portable weld-time meter

has a 5-ampere movement and a dial calibration of 0 to 600 amperes, full scale. In some cases, it is desirable to use a meter movement of other than the 5-ampere range, depending on the maximum kilovolt-ampere rating of the machine and power supply voltage. For example, it has been preferable in some installations to use a 2.5-, 10-, or even a 20-ampere range meter. This selection is determined from the maximum welder transformer primary current divided by the current transformer ratio. The recommended location for mounting of the meter is on the side frame of the machine near the front and above the welder throat or work area. Fig. 4 shows one example of installation. This location provides convenience of accessibility by machine operators.

The third step in preparing the machine is in obtaining the turns ratio characteristic of the welder transformer for the various primary taps. These data in most cases are immediately available from the manufacturer's instruction book. Where the data are not available, the turns ratios are determined by measuring the voltage ratio. The primary voltage divided by the open-circuit secondary voltages for the various primary tap settings is indicative of the welder transformer ratios. To do this, a jumper is placed across the ignitron contactor tube terminals, and voltage applied by a manual breaker or switch. A typical machine data sheet is illustrated in Table I. This is a standard form, which is suitably framed and posted on or near the machine. In practice, turns ratio = voltage ratio.

Making Machine Calibration Records

The machine calibration records of current, time, and weld force are made in the following order:

1. The short-circuit current calibration is made basically as outlined by the Resistance Welders Manufacturers Association (RWMA) Equipment Standards.[1] This means that the secondary short-circuit path is adjusted to a recommended throat opening and depth, and short-circuiting bars used as specified. Normally this work requires solid copper bars of the same diameter as the electrode holder. In the case of a projection-press welder, a short-circuiting bar of recommended length and diameter is placed between the faces of the platens and full head pressure is applied. In this case, the calibration is used for projection work. Where the machine is to be used for spot welding, a calibration is made with the short-circuiting bar solidly clamped in the welder horns, extended to the RWMA recommended throat depth and opening for the size of machine involved. In the making of spot-welder current calibrations where the short-circuiting bar is solidly clamped between the welder horns, the pneumatic pressure should be removed from the machine. Where a machine is used for both spot and projection work, two sets of current calibrations are made for the respective work applications. The mathematics of calculating short-circuit secondary currents from ammeter readings is quite simple:

Secondary current = ammeter reading × current transformer ratio × welder transformer ratio

Table I gives an example of how to set the meter for a required welding or short-circuit current. Current readings are usually taken on every ten points of the heat-control dial setting for the highest primary tap. On the other low and intermediate primary tap settings, readings are taken at three points only; for example, at 40%, 70%, and 100% heat. Full-load and no-load primary supply voltage readings are also observed and recorded on the current calibration chart and throat dimensions are also recorded. Table II shows a typical form used for this purpose.

2. The calibration of the weld-time function is then made before the short-circuiting bar is removed. The primary tap or heat-control setting is reduced to some intermediate or low value for this work. A number of standard cycle-counter instruments may be adapted as a standard for calibration, but the one found most practical is the Westinghouse style *1501174* weld-time meter, which is conveniently connected to the ignitron circuit by means of large, rubber-insulated clip leads. An important reason for this preference is that the instrument may be connected for maintenance testing without interrupting operation of the machine and production. This instrument is shown in Fig. 5. The operation of all timing switches and/or major points of all timing potentiometers is checked. If the error in cycles exceeds the control manufacturer's accuracy tolerance guarantee, the *RC* timing circuit is adjusted or repaired according to requirements. On some control units, a manual adjustment is provided. The weld timer calibration is recorded as shown in Table III.

General Electric Company

Fig. 6. Weld-force gage

3. The short-circuiting bar is then removed and the pressure calibration is made. The weld–no-weld switch should be turned to the "no-weld" position for this work. A force gage is inserted between the electrodes of the machine and head pressure applied. Major points on the pressure gage are checked by adjusting the air regulator and recording force-gage reading in pounds. A preferred type of portable, low-cost gage for this work is the General Electric spring-type force gage, catalog no. *6987801G1*, as shown in Fig. 6. A machine weld-force calibration chart is illustrated in Table III.

Preparing Welding Schedule Charts

A considerable amount of work has been done in the manufacturing laboratory to develop optimum resistance-welding schedules for commercial welding to both mild steel and aluminum. One phase of endeavor in this work has been to keep the machine setting of variables to a minimum for a maximum range of

Table III. Weld Force and Time Calibration for Precision 250-Kva Rocker Arm Spot Welder

Weld-Pressure Calibration		Welder Timer Calibration	
Pressure Gage	Weld Pressure, Pounds	Timing Range Switch	Cycles, ±
10	370	1–30	0
20	980	31–60	1
30	1,490	61–90	3
40	2,050	91–120	5
50	2,630	121–150	7
60	3,180	151–180	9
70	3,720	181–210	16
		211–240	20

To make correction when selecting weld time: Subtract correction in time cycles as indicated from required weld time. Set timer switches to this value.

Table IV. Resistance Spot-Welding Schedules for "Mild Steel" (0.18 Carbon Maximum)

WELDING DATA							INSPECTION & ENGINEERING DATA					
THK COMB. NOTE 1 (IN.)	ELEC-TRODE DIA & SHAPE COMB. SEE BELOW	WELD-ING FORCE LB NOTE 2	WELD TIME CYCLES	WELD-ING CUR AMP NOTE 3	SHORT CKT CUR THRU COPPER NOTE 4	HOLD TIME CYCLES NOTE 5	DIA OF WELD NUGGET OF MULTI-WELDS INCHES NOTE 6	MIN. SPACING IN.	MIN. MULTI-WELD SHEAR STR LB BASED ON STL OF ULT STR OF 45,000 P.S.I.	SINGLE WELD SHEAR STR LB NOTE 7	MIN. EDGE DISTANCE IN.	MIN. OVER-LAP DIS-TANCE	THK COMB. NOTE 1 (IN.)
.015-.015	1 or 2	250	3	7000	8200	10	3/32	1/4	150	150	1/8	1/4	.015-.015
-.036	1 or 2	250	9	7000	9200	10	5/32	1/4	150	200	5/32	5/16	.036
.050	1 or 2	250	9	7000	9200	10	5/32	1/4	150	200	5/32	5/16	.050
.064	1 or 2	250	9	7000	8200	10	5/32	1/4	200	200	5/32	5/16	.064
.036-.036	1 or 2	250	9	7000	8200	10	3/32	1/4	350	500	5/32	5/16	.036-.036
.050	1 or 2	400	14	8400	9900	20	13/64	3/8	550	700	5/32	1/2	.050
.064	1 or 2	400	14	8400	9900	20	13/64	3/8	550	700	1/4	1/2	.064
.078	1 or 2	600	21	9200	10800	20	3/16	1/2	700	1000	1/4	1/2	.078
.094	1 or 2	600	21	9200	10800	20	3/16	1/2	700	1050	1/4	1/2	.094
.109	2	800	28	11400	13400	30	9/32	5/8	900	1250	5/16	5/8	.109
.125	2	800	28	11400	13400	30	1/4	5/8	900	1300	5/16	5/8	.125
.050-.050	1 or 2	400	14	8400	9900	20	3/16	3/8	725	1000	1/4	1/2	.050-.050
.064	1 or 2	600	21	9200	10800	20	1/4	1/2	1000	1500	1/4	1/2	.064
.078	1 or 2	600	21	9200	10800	20	1/4	1/2	1000	1500	1/4	1/2	.078
.094	2	800	28	11400	13400	30	1/4	5/8	1100	1700	5/16	5/8	.094
.109	2	800	28	11400	13400	30	17/64	5/8	1100	1700	5/16	5/8	.109
.125	2 or 3	1050	37	11400	13400	30	17/64	7/8	1700	2200	3/8	3/4	.125
.156	2 or 3	1300	47	14800	17200	30	5/16	7/8	1800	2400	3/8	3/4	.156
.064-.064	1 or 2	600	21	9200	10800	20	9/32	1/2	1100	1800	1/4	1/2	.064-.064
.078	2	800	28	11400	13400	30	1/4	5/8	1800	2300	5/16	5/8	.078
.094	2	800	28	11400	13400	30	15/64	5/8	1700	2200	5/16	5/8	.094
.109	2 or 3	1050	37	11400	13400	30	9/32	7/8	2100	2700	3/8	3/4	.109
.125	2 or 3	1050	37	11400	13400	30	1/4	7/8	2100	2700	3/8	3/4	.125
.156	2 or 3	1300	59	14800	17300	30	5/16	1	2400	3100	7/16	7/8	.156
.188	2 or 3	1500	59	14800	17200	30	1/4	1	2400	3300	7/16	7/8	.188
.250	3	2200	93	19200	22800	45	9/32	1	2900	4100	1/2	1	.250
.078-.078	2	800	28	11400	13400	30	9/32	5/8	1900	3400	5/16	5/8	.078-.078
.094	2 or 3	1050	37	11400	13400	30	17/64	7/8	2800	3700	3/8	3/4	.094
.109	2 or 3	1050	37	11400	13400	30	9/32	7/8	2900	3900	3/8	3/4	.109
.125	2 or 3	1300	47	14800	17200	30	11/32	7/8	2900	3900	3/8	3/4	.125
.156	2 or 3	1500	59	14800	17200	30	5/16	1	2900	3900	7/16	7/8	.156
.188	2 or 3	1500	59	14800	17200	30	11/32	1	3000	3900	7/16	7/8	.188
.250	3	2250	93	19200	22800	45	11/32	1	3000	4500	1/2	1	.250
.094-.094	2 or 3	1050	37	11400	13400	30	9/32	7/8	2700	3800	3/8	3/4	.094-.094
.109	2 or 3	1300	47	14800	17200	30	5/16	7/8	3000	3900	3/8	3/4	.109
.125	2 or 3	1300	47	14800	17200	30	5/16	1	3300	5000	3/8	3/4	.125
.156	2 or 3	1500	59	14800	17200	30	11/32	1	4300	5500	7/16	7/8	.156
.188	3	2250	93	19200	22800	45	3/8	1	4400	6500	1/2	1	.188
.250	3 or 4	2700	130	22200	26100	45	3/8	1-1/8	5000	7900	5/8	1-1/4	.250
.109-.109	2 or 3	1300	47	14800	17200	30	5/16	7/8	2900	4000	3/8	3/4	.109-.109
.125	2 or 3	1500	59	14800	17200	30	3/8	1	3300	5900	7/16	7/8	.125
.156	2 or 3	2200	93	19200	22800	30	7/16	1	4800	6800	7/16	7/8	.156
.188	3	2200	93	19200	22800	45	7/16	1	4800	7900	1/2	1	.188
.250	3 or 4	2700	130	22200	26100	45	15/32	1-1/8	5100	8200	5/8	1-1/4	.250
.125-.125	2 or 3	1500	59	14800	17200	30	5/16	1	4000	4800	7/16	7/8	.125-.125
.156	3	2200	93	19200	22800	45	15/32	1	5000	7300	1/2	1	.156
.188	3	2300	93	19200	22800	45	7/16	1	5100	7900	1/2	1	.188
.250	3 or 4	2700	130	22200	26100	45	5/8	1-1/8	6300	9300	5/8	1-1/4	.250
.156-.156	3	2200	93	19200	22800	45	1/2	1	6400	9400	1/2	1	.156-.156
.188	3 or 4	2700	130	22200	26100	45	19/32	1-1/8	7200	10500	5/8	1-1/4	.188
.250	3 or 4	2700	130	22200	26100	45	19/32	1-1/8	7400	11800	5/8	1-1/4	.250
.188-.188	3 or 4	2700	130	22200	26100	45	1/2	1-1/8	8000	13600	5/8	1-1/4	.188-.188
.250	4	3700	230	28000	33000	90	5/8	1-3/8	11700	14000	3/4	1-1/2	.250
.250-.250	4	3700	153	28000	33000	90	3/4	4	15000	15000	3/4	1-1/2	.250 .250
.250-.250	4	3700	230	28000	33000	90	5/8	1-3/8	13500	15000	3/4	1-1/2	.250-.250

COMBINATION 1: .492 D-2 IN. SPHERICAL RADIUS FACE
.492 D-FLAT

COMBINATION 2: 5/8 D-2 IN. SPHERICAL RADIUS FACE
5/8 D-FLAT

COMBINATION 3: 7/8 D-6 IN. SPHERICAL RADIUS FACE
7/8 D-FLAT

COMBINATION 4: 1-1/2 D-6 IN. SPHERICAL RADIUS FACE
1-1/2 D-FLAT

Explanatory Notes Pertaining to Chart on Spotwelding Schedules

Note 1: On dissimilar thickness combinations, the average thickness determines the weld schedules. These average thicknesses are established to the nearest similar thickness combination. Example, average thickness of 0.050 to 0.094 is 0.072. Average thickness is taken as 0.078. For similar thicknesses not listed, use nearest thickness combination.

Note 2: Welding Force is the net electrode force measured in pounds exerted on the work.

Note 3: Welding Current is the secondary current during the actual welding cycle. It is not the same as Short-Circuit Current. Welding Current variation during welding should be held to ±5%.

Note 4: Short-Circuit Current through copper is used as an aid in setting the welding machine to the schedules. If this value is obtained while firing the machine through copper, it will give approximately the correct welding current. Final adjustments should be made on the work.

Note 5: Hold time is the time required to solidify the weld under pressure after the welding current has been terminated. This time is set on the welder control.

Note 6: The diameter of the weld nugget is the basis for determining the weld strength. The weld diameter should not be confused with the imprint of the electrodes on the work.

Note 7: Shear strengths are obtained by pulling the overlapped weld joints to destruction in a standard testing machine with the specimens unconfined. These are generally known as Tension-Shear tests. These particular shear strengths are obtained with widths equal to 4 times the minimum edge distance.

Table V. Mild Steel Schedule Chart, Precision Rocker-Arm 250-Kva Spot Welder

Two Thicknesses of	Welding Current, Amperes	Meter Reading for Welding Current	Short-Circuit Current Through Copper, Amperes	Meter Reading Through Copper	Meter Range Switch	Heat Control Dial, Per Cent Approximate	Tap Setting	Welding Force, Pounds	Gage Pressure, Pounds	Weld Time, Cycles	Hold Time, Cycles	Electrode Diameter and Shape Combination*
0.036	7,000	230	8,200	270	1	35	Lo-1	250	5	9	10	A or B
0.050	8,400	280	9,900	330	1	40	Lo-1	400	8	14	20	A or B
0.064	9,200	310	10,800	360	1	45	Lo-1	600	12	21	20	A or B
0.078	11,400	380	13,400	450	1	50	Lo-1	800	16	28	30	A or B
0.094	11,400	380	13,400	450	2	30	Hi-4	1,050	20	37	30	C
0.109	14,600	490	17,200	575	2	35	Hi-4	1,300	25	47	30	C
0.125	14,600	490	17,200	575	2	35	Hi-4	1,500	30	59	30	C
0.156	19,200	330	22,600	380	3	40	Hi-4	2,200	42	93	45	C
0.188	22,200	370	26,100	435	3	47	Hi-4	2,700	52	130	45	C or D
0.250	28,000	470	33,000	550	3	55	Hi-4	3,700	70	153	90	D, 2 single welds only.
0.250	28,000	470	33,000	550	3	55	Hi-4	3,700	70	230	90	Same as above. Use 230-cycle timing on second or more welds

Notes on Welding Schedule:

1. Short-Circuit Current Through Copper or Meter Reading Through Copper.
   These columns are to aid in setting up the machine for the required welding current. If this value is obtained while firing the machine through copper, it will give approximately the correct welding current. Meter readings should be established with one division deflection of the pointer. If there is no pointer deflection, increase the heat per cent until recommended reading is obtained.
2. Welding Current or Meter Reading for Welding Current.
   These are the values that the machine must be set for when the welds are actually being made. If more than one weld is made, the machine must be set to obtain these values for the spot weld which has the most metal in the throat of the machine. Spitting or expulsion of metal should not be tolerated. If this occurs, decrease Heat Control dial setting by one or two points, until spitting or expulsion stops.
3. Approximate Heat Dial Per Cent.
   These values are only approximate and are figured with the throat opening and throat depth of the machine set where they are most likely to be used. They are given to aid in setting up the machine. If the throat opening or throat depth of the machine is increased, the Heat Dial per cent will have to be increased to obtain the correct current during welding. Also, if metal is placed in the throat of the machine during the welding cycle, the Heat Dial per cent will have to be increased to obtain correct current during welding.
4. Electrodes.
   Use the electrodes given in the chart above for the various thicknesses of material. Do no file the electrodes in the machine. If the electrodes become worn, replace them with new or refaced electrodes.
5. Thickness Combinations.
   When two unequal thicknesses are to be welded, use the machine settings for the average thickness. For example, in welding one thickness of 0.125 inch to one thickness of 0.188 inch the average thickness is 0.156 inch, and settings for this thickness are used. See Weld Schedule Chart for other dissimilar thickness combinations.
6. Hold Time.
   The hold time shown for each thickness should be rigidly maintained. Insufficient hold time results in greatly weakened welds.

If something goes wrong with the machine or welding schedule settings cannot be maintained, notify the supervisor at once.

* Electrode diameter and shape combinations:
A—1/2-inch-diameter 2-inch-radius face and 1/2-inch-diameter flat
B—5/8-inch-diameter 2-inch-radius face and 5/8-inch-diameter flat
C—One 7/8-inch-diameter flat and one 7/8-inch-diameter 6-inch-radius face
D—One 1¼-inch-diameter flat and one 1¼-inch-diameter 6-inch-radius face

work. In the case of mild steel schedules, a company-wide process specification has been issued and the schedule is included as an integral part. Thus, the requirements for the variables are determined. Typical data, as developed and accepted for the company-wide resistance welding of mild steel, are given in Table IV.

The first step in the preparation of a schedule chart is to compare the machine output in terms of maximum force, time, and current with these requirements for the widest range of work possible at these maximum values. To illustrate how this is done, assume that a schedule chart for mild steel is desired. Also for illustration, assume that the machine under consideration is a 250-kva Rocker-arm type, whose calibration of variables is as given in Tables II and III. The maximum short-circuit current output of the machine is 47,300 amperes, which is well above the requirement of welding two pieces of 0.250-inch mild steel (33,000 amperes per Table IV). It is also noted from the schedule that a weld force of 3,700 pounds is required for welding two pieces of 0.250-inch mild steel. The force system on this particular machine is then adequate for this range of work. The next reference is made to the weld-time requirements for the 0.250-inch combination thickness material and is noted to be 230 cycles. A check with the weld-time function on the calibration chart shows that this requirement can be met. Thus, the machine is capable of welding a range of work up to two thicknesses of 0.250-inch material.

Table VI. Resistance Welding Schedules

Variable	Schedule	
	Single-Force	Dual-Force
Current, amperes		
Weld	33,000	33,000
Starting	10,000	10,000
Final	20,000	10,000
Time, cycles		
Rise	5	5
Taper	16	8
Total	32	25
Cycles of weld before taper	16	16
Force, pounds		
Weld	750	750
Forge		1,500
Forge time delay, cycles		14

A typical mild steel schedule chart is shown in Table V. All machine settings

Fig. 7. Rapid time selector box and current meter

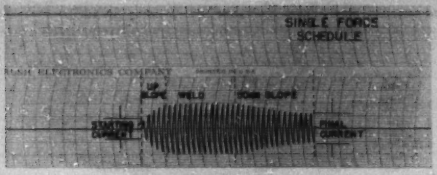


Fig. 8. Single-force schedule

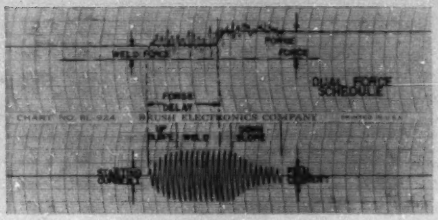


Fig. 9. Dual-force schedule

Brush Development Company

Fig. 10. Portable recording equipment

were derived from the welding variable requirements as outlined in the process specification schedule. Specifically, the current ammeter settings for "short circuit" and for "welding current" are calculated as outlined in Table I.

It should be noted that the purpose of including meter readings for short circuit or "through copper" is to enable the welding operator to adjust the machine heat-control setting to a close value for welding current without actually firing the welder through steel parts. For setup purposes, the operator merely places a piece of 1/8-inch-thick copper between the welding electrodes and adjusts the heat control until he obtains the current meter deflection as indicated on the chart for "copper." This value automatically drops to the "meter reading for welding current" value, when welding current is passed through the actual steel work. The only operator requirement, then, is to lower the back-stop on the ammeter to the setting for "welding current" and observe the meter deflection. No further adjustment of heat control is ordinarily required.

It is also noted that a meter range switch is a desirable feature for selecting the required primary current transformer ratio. This feature permits the selection of the most desirable or uppermost scale deflection.

In making up the schedule chart, the remaining machine settings for weld force, weld time, and recommended electrode combinations are determined from the schedule requirements, Table IV, and the corresponding machine settings from the calibration charts, Table III.

A convenient time-selector switch is frequently used for the rapid selection of specific weld-time machine settings. The switch is calibrated to select and impose predetermined values of resistance R in the welding control RC timing circuit. Fig. 7 shows a mounting of the time selector directly below the weld current meter. In using this time selector, the convenience and rapid setup affords an obvious advantage in time saving over the regular setting of control time dials and/or switches.

Resistance-Welding Instrumentation for Aluminum Work

All procedures for machine preparation and calibration as outlined for mild steel spot welding are identical for the single-phase spot welding of aluminum. Other recommended resistance-welding schedules in use for this work involve additional requirements for the following supplementary instrumentation:

1. *Weld current measurement.* In the in-

C. B. Smith Company

Fig. 11 (left). Strain-gage installation

C. B. Smith Company

Fig. 12 (right). Surface resistance analyzer

Table VII. Recommended Single-Phase Resistance-Spot-Welding Schedules for Similar and

Thickness Combination, Inches	Electrode Combination	Welding Force, Pounds (Note 1)	Up-Slope Time, Cycles (Note 2)	Up-Slope Preheat Current, Amperes (30% of Weld Current) (Note 3)	Weld Heat Time, Cycles (Note 4)	Weld Current, Amperes (Note 5)	Cycles of Weld Before Down-Slope Time (Up Slope plus Weld Time) (Note 6)	Down-Slope Postheat Time, Cycles (Note 7)	Final Down-Slope Current, Amperes (60% of Weld Current) (Note 8)	52SH34 to			
										52SH34	61ST6	24ST3	3SH14
0.032–0.032	A	500	4	7,500	8	25,000	12	10	15,000	400	400	450	320
0.032–0.064	A	600	4	8,500	10	28,000	14	12	17,000	650	675	650	550
0.032–0.091	C or F	750	4	10,000	12	33,000	16	16	20,000	1,050	925	950	600
0.032–0.125	F	1,000	4	11,500	12	38,000	16	16	23,000	1,450	1,150	1,300	1,050
0.064–0.032	A	600	4	8,500	10	28,000	14	12	17,000	650	675	650	550
0.064–0.064	A or D	750	4	10,000	12	33,000	16	16	20,000	1,125	1,050	1,175	1,000 (NM)
0.064–0.091	F	1,000	4	11,500	12	38,000	16	16	23,000	1,450 (D)	1,450 (D)	1,650 (E)	1,400 (F)
0.064–0.125	F	1,200	4	12,000	12	40,000	16	18	24,000	1,850	1,650	1,750 (E)	1,700
0.091–0.032	B or E	750	4	10,000	12	33,000	16	16	20,000	1,050	925	950	600
0.091–0.064	E	1,000	4	11,500	12	38,000	16	16	23,000	1,450	1,700	1,650	1,250 (F)
0.091–0.091	D	1,200	4	12,000	12	40,000	16	18	24,000	1,800	2,000	1,800 (E)	1,490 (F)
0.091–0.125	F	1,600	4	13,500	14	45,000	18	18	27,000	2,000	2,150	2,700	1,770
0.125–0.032	E	1,000	4	11,500	12	38,000	16	16	23,000	1,450	1,150	1,300	1,050
0.125–0.064	E	1,200	4	12,000	12	40,000	16	18	24,000	1,850	1,850	1,750	1,350 (F)
0.125–0.091	E	1,600	4	13,500	14	45,000	18	18	27,000	2,000	2,400	2,200 (NM)	1,445 (F)
0.125–0.125	D	2,000	4	15,000	15	50,000	19	20	30,000	2,250	1,900	3,500	2,000 (NM)
0.125 24ST3–0.125 24ST3	D	3,000	4	16,500	15	55,000	19	20	33,000				

Note 1. Welding force is the net electrode force measured in pounds exerted on the materials being welded.
Note 2. Up-slope time is the total time for the up-slope current to rise from 30% of the weld current to full value of the weld current. This is 4 cycles for all thickness combinations between 0.032 and 0.125 as shown.
Note 3. Up-slope current is the impulse of current which flows from the end of squeeze time to the beginning of weld time.
Note 4. Weld heat time is the time during which welding current flows through the materials being welded.
Note 5. Welding current is the secondary current during the actual welding cycle.
Note 6. "Cycles of weld before down slope" is the dial setting calibrated in cycles and is the sum of the up-slope time plus the weld heat time cycles.
Note 7. Down-slope time is the total time during which postheat or taper current flows.
Note 8. Down-slope current is the impulse of current which flows from the end of the weld time cycle tapering down to 60% of the weld current.
Note 9. Tensile shear strengths are obtained by pulling the overlapped spot-weld joints to destruction in a standard testing machine with the specimens unconfined. These particular shear strengths are obtained with specimen widths equal to recommended spot spacings. The above strengths are the average of 100 multispot welds. The letter in parentheses adjacent to strength values indicates the electrode combination required. For example, the combination of 0.064 *61ST6* to 0.091 *24ST3* shows a tensile shear designation of 1,550(E). By this is meant that this combination has a tensile shear strength of 1,550 pounds. Also, referring to the electrode combination reference for (E), it can be said that a 7/8-inch-diameter, 8-inch-spherical-radius-face electrode is used against the 0.064 thickness of *61ST6* and a 7/8-inch-diameter flat electrode is used against the 0.091 *24ST3* material. Similarly, with 0.064 *52SH34* welded to 0.091 *3SH14* material, the electrode combination is (F). Thus, a 7/8-inch-diameter flat electrode is used against the 0.064 *52S* material and a 7/8-inch-diameter, 8-inch-spherical-radius-face electrode is used against the 0.091 *3S* material. The designation (NM) adjacent to strength values indicates that this particular combination is not a military quality combination; i.e., the minimum weld nugget penetration required in Military Specification *6860* cannot be met with these schedules.

General Note. Minimum squeeze and hold time should be 15 cycles.

strumentation of weld current where a modulated current waveform is used (i.e., where the current is initially applied at some low value, increases to a nominal or welding value, and then tapers off to a final amplitude), a medium of measurement is required which will register the relative current amplitude values in a chronological manner. A brush amplifier and oscillograph has proved to be a practical instrument for this application.

For setup purposes, the weld current meter provides the necessary medium for machine setting to the nominal or maximum weld current. This measurement is initially made with a constant-amplitude current waveform (no up-slope or down-slope) using a minimum 15-cycle timing.

In this setup of the weld current meter to the required aluminum schedule current, it is noted that only one current setting is necessary. The weld current for aluminum has been found to be virtually the same as that of the short-circuit current values; i.e., when the welder is fired through a piece of copper strap between the welding electrodes for setup purposes, this same meter reading is usually obtained when the welder is fired through the actual aluminum work. Thus the machine schedule chart has but one current setting for the material thickness involved.

The brush oscillograph is then adjusted for an average pen deflection of 10 millimeters, or divisions on each side of center-zero. This adjustment is arbitrarily made for the scheduled weld current, or as determined in using the weld current meter reading.

The up-slope and down-slope current amplitude is then adjusted on the machine control and the waveform observed on the oscillograph. The amplitudes of starting and final current are read as a per cent of the maximum, or 10-millimeter pen deflection. A typical brush recording of a modulated current waveform is shown in Fig. 8. This record represents current readings for the spot welding of 0.064-inch aluminum type *52SH34*, using a single-weld-force system.

2. *Instrumentation of weld force* (*S*). In the spot welding of aluminum using weld schedules which are developed for single-force systems, the same medium of measurement is adequate as that for mild steel spot welding. This involves the setting of an air gage on the machine to render the desired schedule weld force.

Where aluminum work is done with dual-force schedules, a medium for measuring the actual "forge time delay" is required. This is true because in the use of dual force or forge pressure, the advantage of maximum grain refinement in the weld will only be obtained when the forge pressure is applied in the exact chronological sequence on the welding cycle. Specifically, a desirable combination of current waveform modula-

Average Tensile Shear Strengths (Note 9)												Inspection and Engineering Data: Recommended Minimum			
61ST6 to				3SH14 to				24ST3 to				Edge Distance†	Overlap‡	Spacing§	Thickness Combination, Inches
61ST6	24ST3	3SH14	52SH34	3SH14	52SH34	61ST6	24ST3	24ST3	3SH14	52SH34	61ST6				
400	400	310	400	300	320	310	300	400	300	450	400	1/4	1/2	3/4	0.032–0.032
650	650	525	675	550	550	525	625 (NM)	700	625 (NM)	650	650	1/4	1/2	3/4	0.032–0.063
1,050	900	600	925	850	600	575	675 (NM)	1,150	675 (NM)	700	700	5/16	5/8	1	0.032–0.091
1,450	1,100	800	1,150	1,085	1,050	800	725	1,600	725	725	725	5/16	5/8	1	0.032–0.125
650	650	525	675	550	550	525	625 (NM)	700	625 (NM)	650	650	1/4	1/2	3/4	0.063–0.032
1,050	1,100	900	1,050	750	1,050 (NM)	900	850 (NM)	1,150	870 (NM)	1,175	1,100	5/16	5/8	1	0.063–0.063
1,450 (D)	1,550 (E)	1,450 (F)	1,700	1,085	1,250 (E)	1,900 (E)	1,150 (NM)	1,600	1,350 (NM)	1,650	1,550	3/8	3/4	1	0.063–0.091
1,850	1,675	1,550 (NM)	1,850	1,200	1,350 (E)	1,900 (E)	1,300 (NM)	1,950	1,450	1,750	1,800	3/8	3/4	1	0.063–0.125
1,050	900	600	925	850	600	575	675	1,150	675 (NM)	700	700	5/16	5/8	1	0.091–0.032
1,450	1,550	1,100 (F)	1,700	1,050	1,400	1,400	1,350	1,600	1,150 (NM)	1,650 (F)	1,550 (F)	3/8	3/4	1	0.091–0.064
1,650	1,825 (E)	1,550 (F)	2,050	1,100	1,500 (E)	1,550 (E)	1,300	2,000	1,290 (NM)	1,800 (F)	1,825 (F)	1/2	1	1	0.091–0.091
1,700	2,400	1,650	2,400	1,250	1,450 (E)	1,425 (E)	1,650 (NM)	2,250	1,450 (NM)	2,190 (NM)	2,090 (E)	1/2	1	1	0.091–0.125
1,450	1,100	800	1,150	1,050	1,050	800	725	1,600	725	725	725	5/16	5/8	1	0.125–0.032
1,650	1,800	1,275 (F)	1,650	1,200	1,700	1,550	1,450	1,950	1,200 (NM)	1,750 (F)	1,675 (F)	5/16	5/8	1	0.125–0.064
1,700 (F)	2,100 (F)	1,425 (F)	2,150	1,260 (F)	1,750 (F)	1,650 (F)	1,500 (NM)	2,250	1,635 (NM)	2,750 (F)	2,400 (F)	1/2	1	1	0.125–0.091
2,250	2,600	1,600 (F)	1,900	1,500	2,000 (NM)	1,600 (E)	1,800 (NM)		1,800 (NM)	3,550	2,600	5/8	1 1/4	1 1/2	0.125–0.125
								3,000				5/8	1 1/4	1 1/2	0.01254 ST3 0.01254 ST3

* Electrode combinations (see diagram, Table IV, column 2):
A—Two 5/8-inch-diameter, 3-inch-spherical-radius face
B—One 5/8-inch-diameter, 3-inch-spherical-radius face and one 5/8-inch-diameter flat
C—One 5/8-inch-diameter flat and one 5/8-inch-diameter, 3-inch-spherical-radius face
D—Two 7/8-inch-diameter, 8-inch-spherical-radius face
E—One 7/8-inch-diameter, 8-inch-spherical-radius face and one 7/8-inch-diameter flat
F—One 7/8-inch-diameter flat and one 7/8-inch-diameter, 8-inch-spherical-radius face
Electrode material is Resistance Welder Manufacturers Association class *I*.

† See diagram, Table IV, column 12.

‡ See diagram, Table IV, column 13.

§ See diagram, Table IV, column 9.

tion and forge pressure requires the control adjustment of forge delay time (after start of the current flow) to render the application of forge pressure just prior to the tapering or down-slope of weld current. The exact time for the forge pressure application is determined to be just 2 cycles (based on 60-cycle per-second timing) before the start of current tapering.

On modern resistance-welding control systems, the setting of forge time delay function does not afford sufficient accuracy to render the foregoing requirement. This is particularly true when considering the certification of welding schedules for military work. The combination of strain gage, brush amplifier, and oscillograph is an acceptable medium for this measurement. Fig. 9 shows an oscillogram of a modulated current wave, and the application of forge force at the correct point in the chronological sequence of the weld cycle. The welding schedule for this particular oscillogram is for 0.064–0.064, type *52SH34* aluminum. A 2-channel oscillograph is used for this purpose. Fig. 10 shows a practical instrumentation console, which combines all components in a portable unit, as required for this work. A strain-gage assembly, mounted on a resistance-welding holder is shown in Fig. 11. This item is commercially available. The specific aluminum schedules which are instrumented in the oscillograms, Figs. 8 and 9, are shown for comparison in Table VI.

3. *Other instrumentation requirements.* The brush oscillograph method of instrumenting the current waveform also gives a check of weld timing. In a production setup, however, the work schedule certification usually does not require the continuous monitoring of the variables just described, and in most cases periodical checks will suffice. However, in the brush equipment as shown in Fig. 10, a provision for remote-control operation of the oscillograph chart is provided whereby each weld can be recorded if desired.

In the spot welding of aluminum, good weld quality depends directly on the degree of cleanliness of the material. Since considerations of dirt and oxide contribute toward an appreciable increase in surface resistance, it has been found necessary in some cases to equip production departments with a facility for measurement to determine acceptable resistance values on the material to be welded. In general, a surface resistance of less than 50 microhms on material to be welded is acceptable. A portable instrument which meets the requirements for this work is the surface resistance analyzer, model *VT11A*, commercially available from the same company which supplies the strain gage pickup device. This instrument is direct reading and has a pneumatic clamp for holding the material during test. The electric probe or contact points for measurement are included as a built-in feature, which eliminates the need for external electric probes. Fig. 12 shows this device.

4. *Aluminum spot-welding schedules and machine schedule charts.* Typical data developed for company-wide use in aluminum work are shown in Table VII. These data represent the schedule requirements when a single-weld-force system is used. The data can be used in conjunction with dual-force operation by reducing the current taper time and final current by 50%. To illustrate, assume a taper time of 16 cycles and a final current of 20,000 amperes as used on single-force machines. For dual-force application, the taper time then would be reduced to 8 cycles and the final current to 10,000 amperes.

Table VIII. Aluminum Spot-Welding Schedule, 250-Kva Taylor-Winfield

Alloy	Thickness Combination Two of	Welding Current, Amperes	Meter Reading for Welding Current	Meter Range Switch	Per Cent Heat Dial, Approximate	Primary Tap Setting	Starting Current, Amperes (Note 1)	Final Current, Amperes (Note 2)	Start Current Dial, Approximate (Note 1)	Final Current Dial, Approximate (Note 2)	Current Rise Time, Cycles	Current Taper Time, Cycles	Total Weld Time, Cycles	Cycles of Weld Before Taper	Weld Force, Pounds (Note 3)	Pressure Gage at Note 3	Forge Time Delay	Electrode Combination*
3SH14	0.004	33,000	3.95	3	85%	Hi-2	9,900	9,000	20	30	4	8	24	16	750	13	16	A or B
52SH34																		
61ST6	0.004	33,000	3.95	3	85%	Hi-2	9,900	9,900	20	30	4	8	24	16	1,500	38	16	A or B
24ST3																		
3SH14	0.091	40,000	7.0	3	85%	Hi-8	12,000	12,000	20	30	4	9	25	16	1,000	22	16	B
52SH34																		
61ST6	0.091	40,000	7.0	3	85%	Hi-8	12,000	12,000	20	30	4	9	25	16	2,200	60	16	B
24ST3																		
3SH14	0.125	50,000	8.75	3	95%	Hi-8	15,000	15,000	25	35	4	10	29	19	2,200	60	19	B
52SH34																		
61ST6																		

Note 1. Starting currents are adjusted and readings taken on the brush. Oscillogram should indicate the "Start" point of the current trace as 30% of maximum deflection as that obtained for weld current. Dial settings for starting current are approximate.

Note 2. Final currents are adjusted and readings taken on the brush. Oscillogram should indicate the "Final" point of the current trace as 30% of maximum deflection as that obtained for weld current. Dial settings for final current adjustment are approximate.

Note 3. Set the retired stroke gage at 60 and the down-stroke gage at 100.

* Recommended electrode combinations:
A—7/8-inch diameter, top and bottom, 3-inch radius, PDS 9438-1 material (average hold time, 60 cycles).
B—7/8-inch diameter, top and bottom, 8-inch radius, PDS 9438-1 material.

Table VIII shows a sample machine schedule chart for spot welding of aluminum. This chart covers the range of work within the confines of machine capacity and production materials. With the exception of current wave-modulation measurement, the steps in preparation of aluminum schedule charts are virtually the same as those in making charts for spot welding of mild steel.

Miscellaneous Applications

Seam Welding

The use of instrumentation for seam welding presents more potential advantage than with other forms of resistance welding because of relatively higher duty cycle, and any variation can be quickly detected. Basic techniques used are virtually the same. The requirements for seam welding are similar to those for spot welding except for weld current considerations. The duty-cycle pattern of current (i.e., on-time and off-time×number of pulses) usually dictates the need for a weld current meter with a movement with a greater dampening characteristic. The standard thermal-type radio-frequency ammeter has proved satisfactory for shop use. The meter movement is encased in the same metal enclosure as the one used for mild steel and aluminum (Fig. 3) and is equipped with the standard pointer back-stop for adjustment.

Higher Speed Welds (Less Than 10 Cycles)

Where applications require short weld time, which is insufficient for normal deflection of the current meter, the following sequence is used:

1. The proper machine settings to render acceptable work quality are determined by empirical, or other, means.

2. For the purpose of job records, the true current is then measured with the primary meter, using an extended weld time (15 cycles minimum). Short-circuit values with "current-through-copper" should be used as the recorded schedule machine setting.

3. The weld time is then reduced to the required value for welding and meter reading noted. If meter deflection is not the same as with the 15-cycle timing, the meter is adjusted for the lower or indicated value. Thus, the weld current meter serves as a monitor in this case and the indicated values of current do not represent true amplitude.

Effective Use and Application of Instrumentation

A broad responsibility for the effective use of instrumentation and recommended procedures starts with product engineering designs. This is particularly true since manufacturing people can only use and standardize methods within the limitations of actual work requirements. Some important items which engineering designers should consider are

1. The proper selection and specification of material and favorable combinations, where possible.

2. The consideration of joint designs which lend themselves more favorably to the use of standard spot-welding electrodes and simple fixturing.

3. Provision in product designs for adequate overlap of material to maintain the necessary edge distance for location of the weld spot. This is important to obtain optimum weld strengths.

A preliminary consideration of these factors will greatly enable standardization of methods and materially reduce shop problems which would ordinarily arise if the basic requirements of the resistance-welding process are not considered in product designs.

Other responsibilities for an effective instrumentation program as herein outlined are

1. Assignment of optimum welding schedules which permit an acceptable quality work with the minimum equipment and labor requirements.

2. Standardization of manufacturing and instrumentation procedures.

3. The posting and maintenance of welding machine calibration and schedule charts for convenient use by the operator or setup man.

4. Organization of a competent machine maintenance program.

5. The assistance of qualified and competent setup people for this work. Operator training programs help considerably.

Conclusions

1. In general, with the objectives of maximum simplicity of operation, low cost, and acceptable accuracy, the pointer back-stop primary meter is preferred for measurement of basic weld-current values.

2. Where modulated current waveform is used, the Brush-type recorder is recommended.

3. In the use of welding schedules with dual-force systems, a strain-gage measurement and recording brush oscillograph has practical value.

4. Wide economic benefits of major significance in labor and material savings can be realized by the use of instrumentation on the production line.

5. Further economic benefits in manufacturing costs are obtained with standardization of equipment and procedures, which is aided considerably by instrumentation.

6. The machine records of instrumented variables determine equipment performance and serve as a tangible aid in maintenance work. This greatly reduces production downtime in the localizing of machine defects.

7. Standard calibrating and instrumentation procedures which are used in accordance with RWMA standards are valuable in checking performance when a new resistance-welding machine is purchased.

8. Improved product designs with uniform consistency are possible at a lower cost where an effective instrumentation program is used.

References

1. Equipment Standards. *Bulletin no. 16*, Resistance Welders Manufacturers Association, June 1952.

2. A Practical Method for Obtaining Consistent Resistance Welds, J. W. Kehoe. *The Welding Journal*, New York, N. Y., Oct. 1950.

3. Spot Welding Aluminum with Single Phase Equipment, J. W. Kehoe, D. R. McCutcheon. *Ibid.*, Oct. 1954.

4. Recommended Practices for Spot Welding Aluminum and Aluminum Alloys. American Welding Society, New York, N. Y., Oct. 1954.

Electrical Features of the Four Corners Pipe Line

J. J. SONNIER
ASSOCIATE MEMBER AIEE

H. N. SILER
NONMEMBER AIEE

THIS PAPER discusses generally the electric power distribution systems, control schemes, supervisory control, telemetering, and communication means employed on a crude-oil pipe line constructed by The Shell Pipe Line Corporation.

The Four Corners pipe line runs from the San Juan and Paradox basins in the Four Corners area to the refinery area in or near Los Angeles. The line is owned by Continental Pipe Line Company (10%), Gulf Oil Company (20%), Richfield Oil Corporation (10%), Shell Oil Company (25%), Standard Oil Company of California (25%), and The Superior Oil Company (10%). It was designed and constructed and is being operated by Shell Pipe Line Corporation, as agent. The route of the pipe line and station locations are shown in Fig. 1. Initial throughput of the line called for the construction of three main-line pump stations and two pressure-reducing stations. The first delivery of crude oil to the Los Angeles area was made in April 1958.

Electric Power Distribution

Commercial electric power supply was available at all of the locations. The power distribution system for the main-line motors is 3-phase 4,160-volt 60 cycle. The two pressure-reducing stations are supplied with 3-phase 480-volt 60-cycle service. At two of the main-line stations the power companies furnished and owned the substations, while at the third location the customer had to furnish, own, and maintain the required substation. All power from the substation is distributed through an underground conduit network. Fig. 2 shows a typical single-line diagram of power distribution system employed at a main-line station. Fig. 3 shows a typical underground conduit network for distribution of power.

For auxiliary equipment having electric motor drivers, the 4,160-volt service voltage is transformed to 480 volts by customer. Lighting and unit motor-operated valves are supplied by 208/120-volt 4-wire service from a transformer, also furnished by the customer.

Main-Line Pump Motors

All of the main-line centrifugal pump units are driven, through limited end-play spacer couplings, by 900-horsepower 3,600-rpm synchronous-speed squirrel-cage induction motors which have normal starting torque and running characteristics. Each motor is equipped with embedded temperature detector coils and space heaters. At two of the stations the motors are installed on a concrete pad outdoors and have weather-protected enclosures with built-in air filters. At the third station the motors are installed indoors and are of the totally enclosed forced-base ventilated type.

Paper 58-1057, recommended by the AIEE Petroleum Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Petroleum Industry Conference, Dallas, Texas, September 15-17, 1958. Manuscript submitted July 7, 1958; made available for printing December 30, 1958.

J. J. Sonnier and H. N. Siler are with the Shell Pipe Line Corporation, Houston, Tex.

The authors wish to acknowledge the information presented in references 1 through 6 as valuable sources of design features incorporated into this pipe line system.

Local Central Control Cabinets

All locations are provided with a local central control panel. The panels are upright and they incorporate all the necessary pilot lights, push-button stations, and miniature indicating instruments. The pilot lights and push-button stations are grouped by units, and provide control and visual indications of the status of the equipment as follows:

Control

1. Start and stop all units.
2. Discharge pressure set point.
3. Emergency stop.
4. Alarm reset.

Visual Indications

1. On-off main pump units.
2. Opened and Closed unit suction and discharge valves.
3. High station case pressure.
4. High station discharge pressure.
5. Low instrument air pressure.
6. High sump level.
7. Low station battery.

Indicating Instruments

1. Station flow.
2. Station suction pressure.
3. Station case pressure.
4. Station discharge pressure.
5. Gravity.

Device Designations

The device designations are based on the proposed AIEE Standard No. 68.[7] The device numbers used in the schematic drawings are detailed in the Appendix. The schematic drawings illustrated are extracts from construction drawings.

Incoming Line Main Circuit-Breaker Control

The main breaker is located adjacent to the substation. It is an outdoor oil type, rated at 7.5 kv, 1,200 amperes, with 250-megavolt-ampere interrupting capacity. The breaker is equipped with motor-wound spring closing mechanism and 48-volt d-c shunt trip. The breaker is arranged to close automatically unless locked out by device number 51 (overcurrent) or emergency stop, when voltage and phase sequence are normal.

Fig. 4 shows the control schematic used for the control of the main circuit breaker. The breaker is equipped with the following protective relays and devices:

Device No.	Function
47	phase reversal and undervoltage
51	overcurrent
52 UV	d-c undervoltage

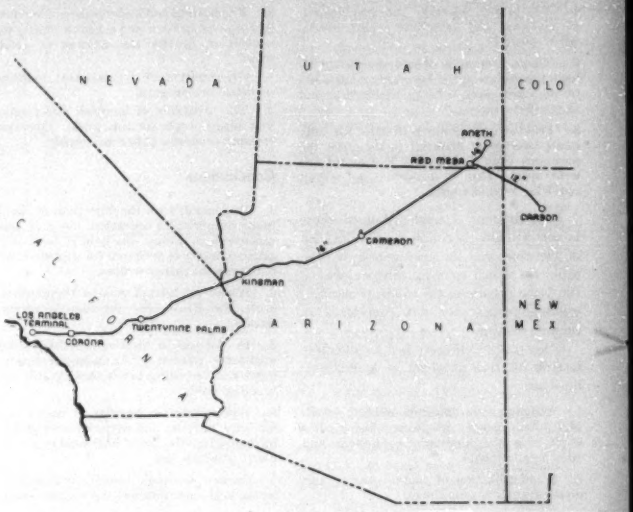


Fig. 1. Pipe line route

The main breaker is locally controlled from three locations: 1. control switch mounted on cubicle 1 of the indoor switchgear assembly, 2. push-button station installed outside the fence enclosure of the breaker, and 3. conventional control switch furnished in the control cabinet of the breaker. Also, the controlling point on the western end of the line by means of supervisory control (to be discussed later) has supervision of each station main breaker.

Main-Line Motor Control and Protective Devices

The power companies' transmission systems dictated the use of reduced voltage starting at all locations for the main pump motor drivers. The reactor method of limiting starting current inrush was satisfactory on all power transmission systems. The motor starters consist of three standard indoor metal-clad switchgear cubicles. Two cubicles house the starting and running circuit breakers. A third cubicle houses the motor-starting reactor. The circuit breakers are of the draw-out, air-break type, rated 5 kv, 1,200 amperes continuous, with interrupting capacity of 150 megavolt-amperes, rectified a-c closing, and 48-volt d-c shunt trip.

Adjacent to each pump unit is a cabinet housing the mechanical protective devices, which are all of the mechanical mercury switch type. The cabinet has a push button installed as an integral part.

The unit protective functions are segregated into four categories: 1. electrical nonlockout, 2. electrical lockout, 3. mechanical nonlockout, and 4. mechanical lockout; see Table I.

Unit Starting and Stopping Sequence

Fig. 5 shows the unit control scheme. Suction and discharge valves must be closed to start the unit. Pressing the "start" push button energizes starting sequence relay 204 and over-all sequence check timer 248. The starting relay closes the control circuit to the suction valve "open" contactor, and energizes the solenoid valve supplying air to the pump vent pneumatic bleeder valve. When the suction valve is 50% opened, the pump vent solenoid valve is de-energized. The fully opened position of the suction valve actuates the control circuit of the starting circuit breaker, causing it to close, and connecting the motor through the reactor to the switchgear busses. The starting breaker closing energizes start-timing relay 219. After preset starting time, the "run" breaker is closed and the start breaker is tripped. The discharge valve "open" contactor is energized by closing the run circuit breaker.

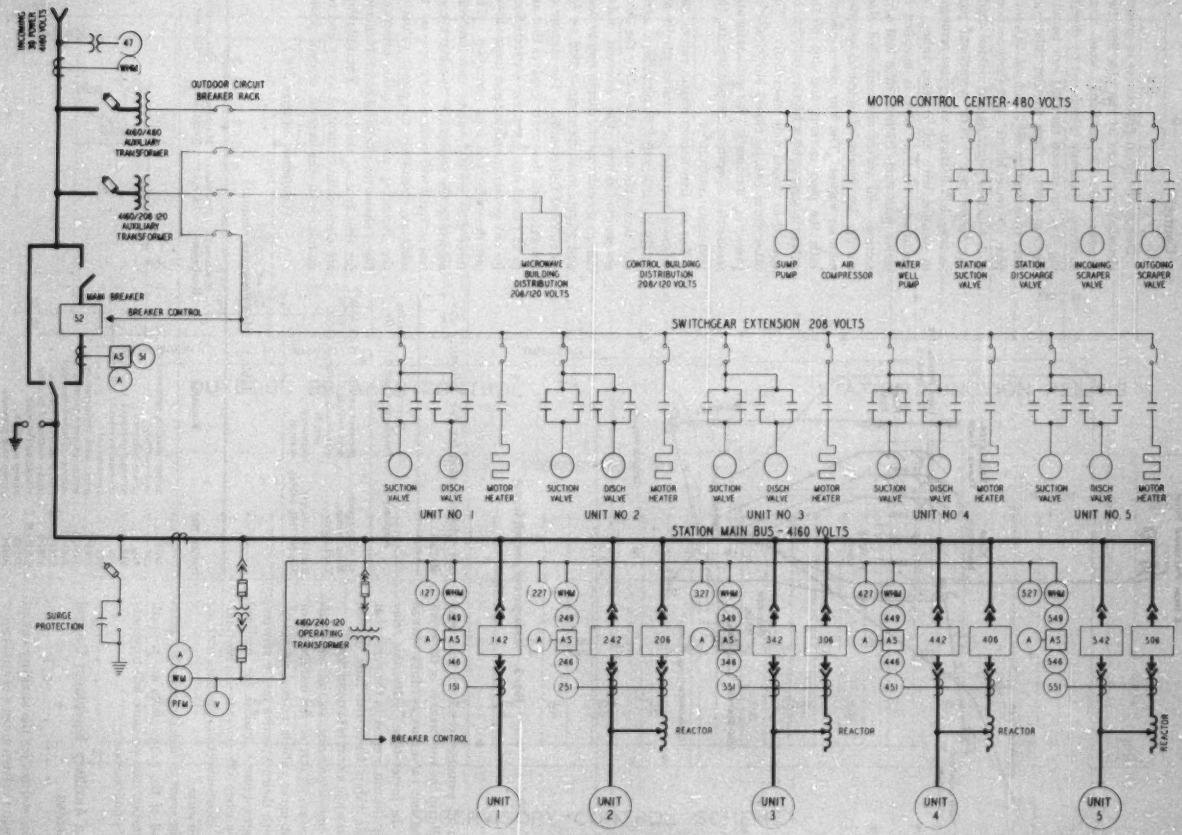


Fig. 2. Single-line diagram

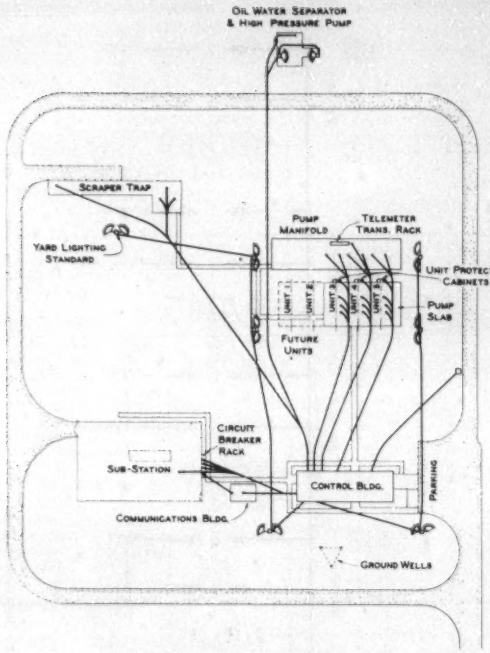


Fig. 3. Underground conduit layout

Over-all sequence check timing relay 248 will trip the run breaker in the event that the unit valves are not fully opened before the preset time. Also, this relay will trip the run breaker in the event that the unit valves leave the fully opened position while a unit is in operation.

All unit-sequencing relays and unit valve-reversing starters are installed in or on switchgear cubicle extensions associated with their respective motor starters. This arrangement makes for unitized motor starters, thereby reducing the number of interconnecting conduits and associated wiring.

In the event of high outgoing line pressure, all of the pump units are tripped. High station case pressure will trip units 1 and 2 simultaneously.

Station Auxiliary Equipment Control Schemes

Motor starters for the auxiliary equipment motor drivers are in a central motor control center. Fig. 6 illustrates the control schemes used for the following auxiliaries:

Auxiliary	Type of Automatic Control	Device No.
Air compressor	pressure switch	63 AIM
Sump pump	float switch	71 SUM
Yard lights	astronomical time switch	42 YL
Battery charger	regulated	
Scraper trap valves	manual	

Pressure-Reducing Station Electric Control

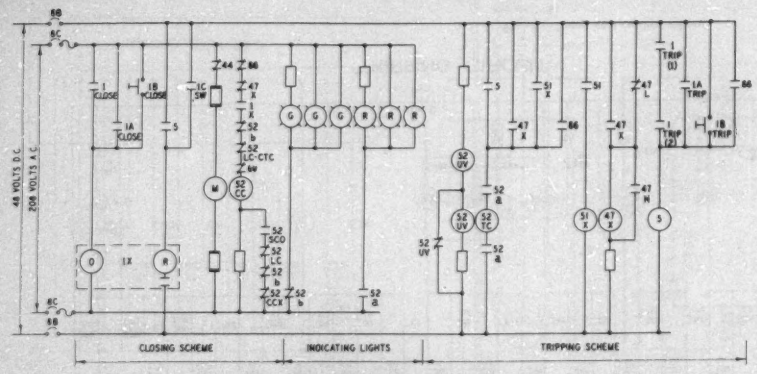
Due to the great differences in ground elevation traversed by the pipe line, two pressure-reducing stations were constructed. Pressure is reduced by diverting the line stream through four parallel friction tubes. Each friction tube consists of a section of 2-inch or $2^1/_2$-inch pipe, 65 feet in length, with three sections of 3/4-inch or 1-inch tubing welded inside the larger pipe. Four additional pipe sections less friction tubing are arranged in parallel with the friction tubes. A pneumatic control valve is installed in series with each friction tube and pipe section. All valves are controlled electrically by means of solenoid-operated pilot air valves. The four valves in series with the pipe sections and one of the valves in series with a friction tube can be opened or closed only electrically. The other three can be closed electrically or placed in "automatic" for automatic and sequential operation responsive to the pneumatic controls.

For the five nonautomatic valves, the solenoid-operated pilot valve admits air to the control valve diaphragm to open and dumps the air to close. For the three automatic valves, the solenoid valve dumps the air from the diaphragm to close and admits instrument-controlled air for partial and full opening. The number of nonautomatic flow control valves that are to be placed into operation is determined for any given throughput and those valves are manually put into operation. The three flow control valves that are connected to the pneumatic automatic control system will sequentially operate to maintain constantly the preset values of the incoming and outgoing pressures. Fig. 7 shows the electrical control schemes for this type of station. The following are electrically controlled locally and remotely from the central control point:

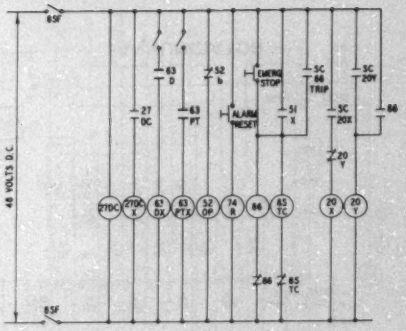
(*Continued on page 218.*)

Table I. Unit Protective Functions

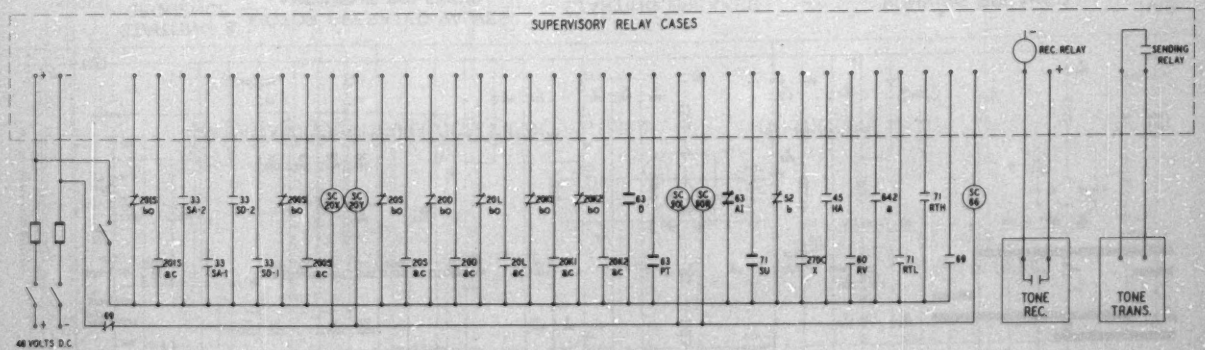
Device No.	Function	Electrical Nonlockout	Electrical Lockout	Mechanical Nonlockout	Mechanical Lockout
226M	motor-winding temperature	X			
227	undervoltage	X			
246	phase balance		X		
249	thermal overload, long time	X			
249	thermal overload, instantaneous		X		
251	overcurrent		X		
226PC	pump case temperature			X	
238M1	motor inboard bearing				X
238M2	motor outboard bearing				X
238P1	pump inboard bearing				X
238P2	pump outboard bearing				X
239PVB	vibration detector				X
239SF	pump seal leak				X
263PC	pump case pressure			X	
263S	pump suction pressure			X	



OUTDOOR BREAKER CONTROL



STATION FUNCTION RELAYS



SUPERVISORY CONTROL SCHEME

Fig. 4. Outdoor breaker control and station functions schemes

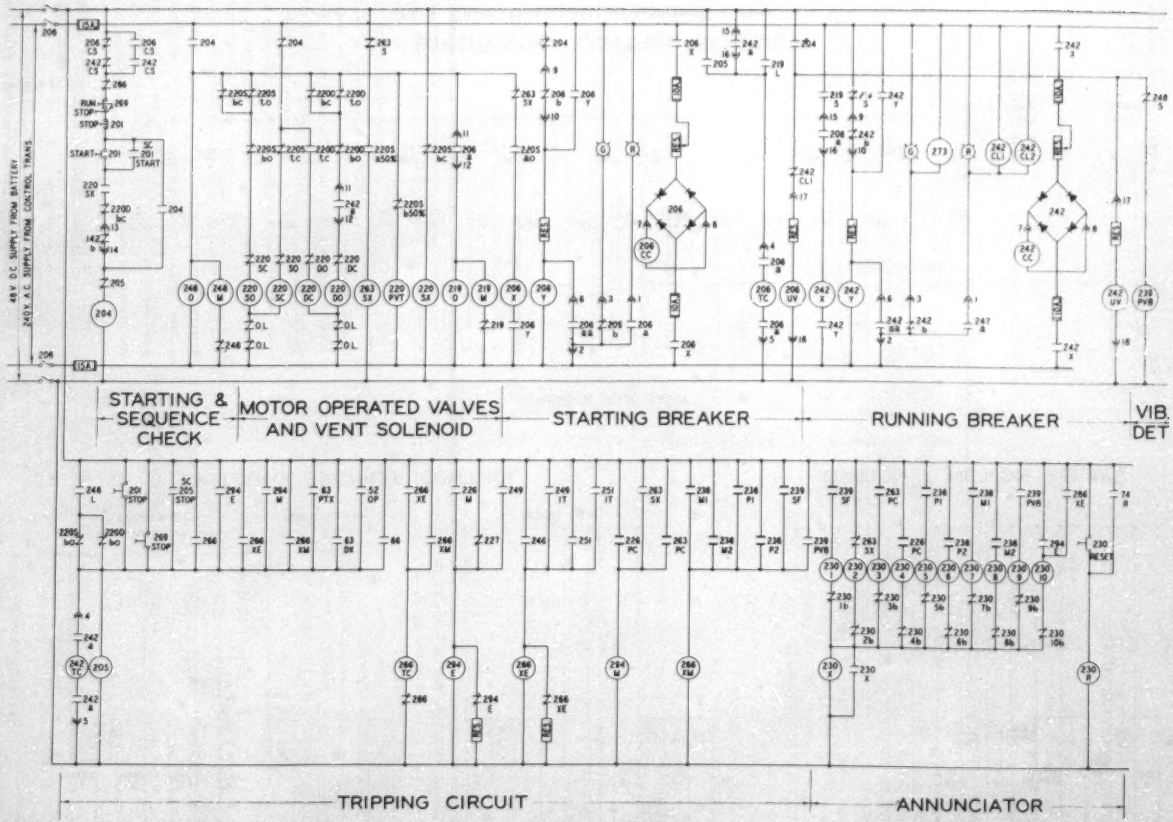


Fig. 5. Pump unit control scheme

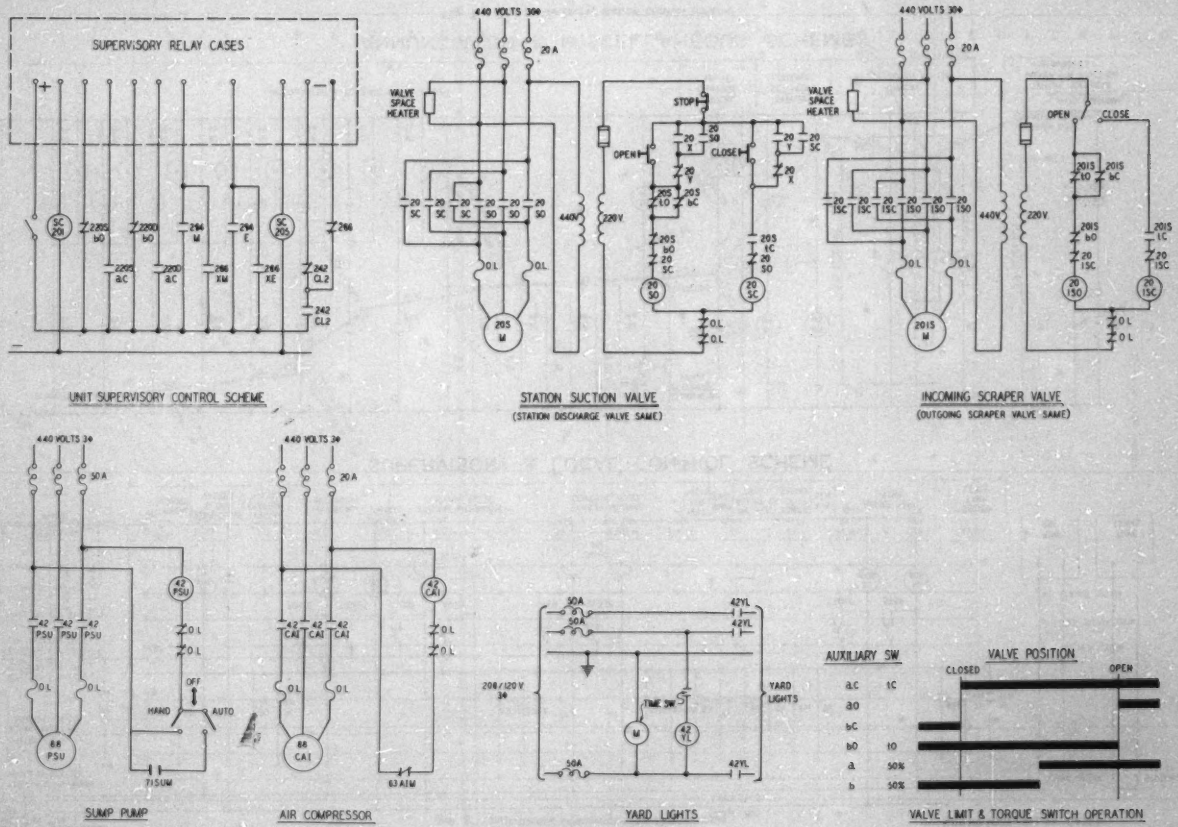
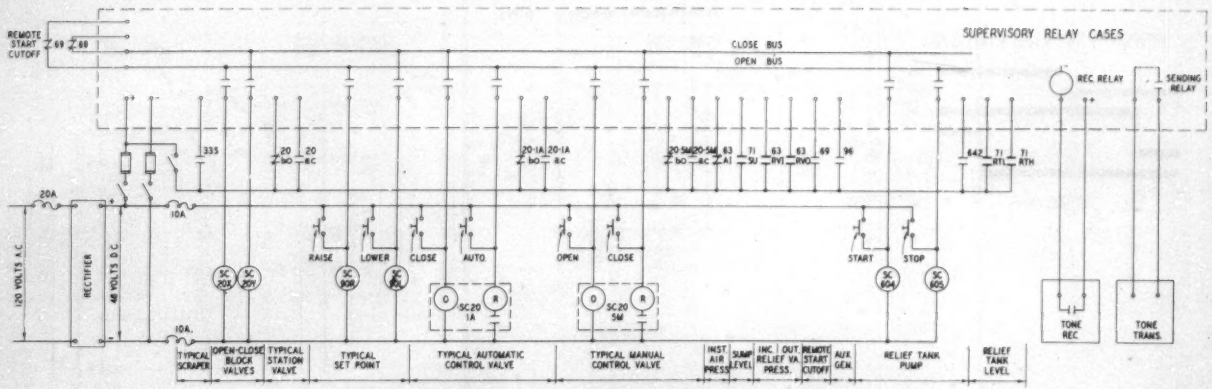
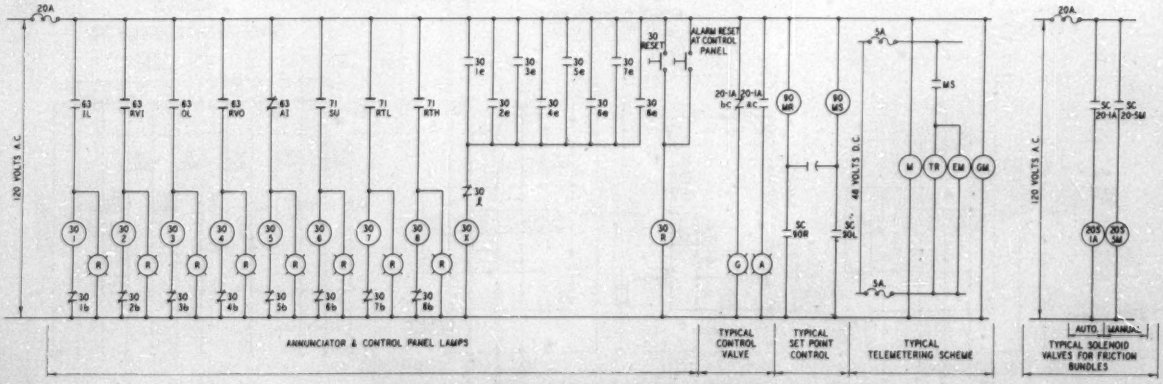


Fig. 6. Auxiliaries control schemes



SUPERVISORY & LOCAL CONTROL SCHEME



ANNUNCIATOR & MISCELLANEOUS SCHEMES

Fig. 7. Pressure-reducing station control scheme

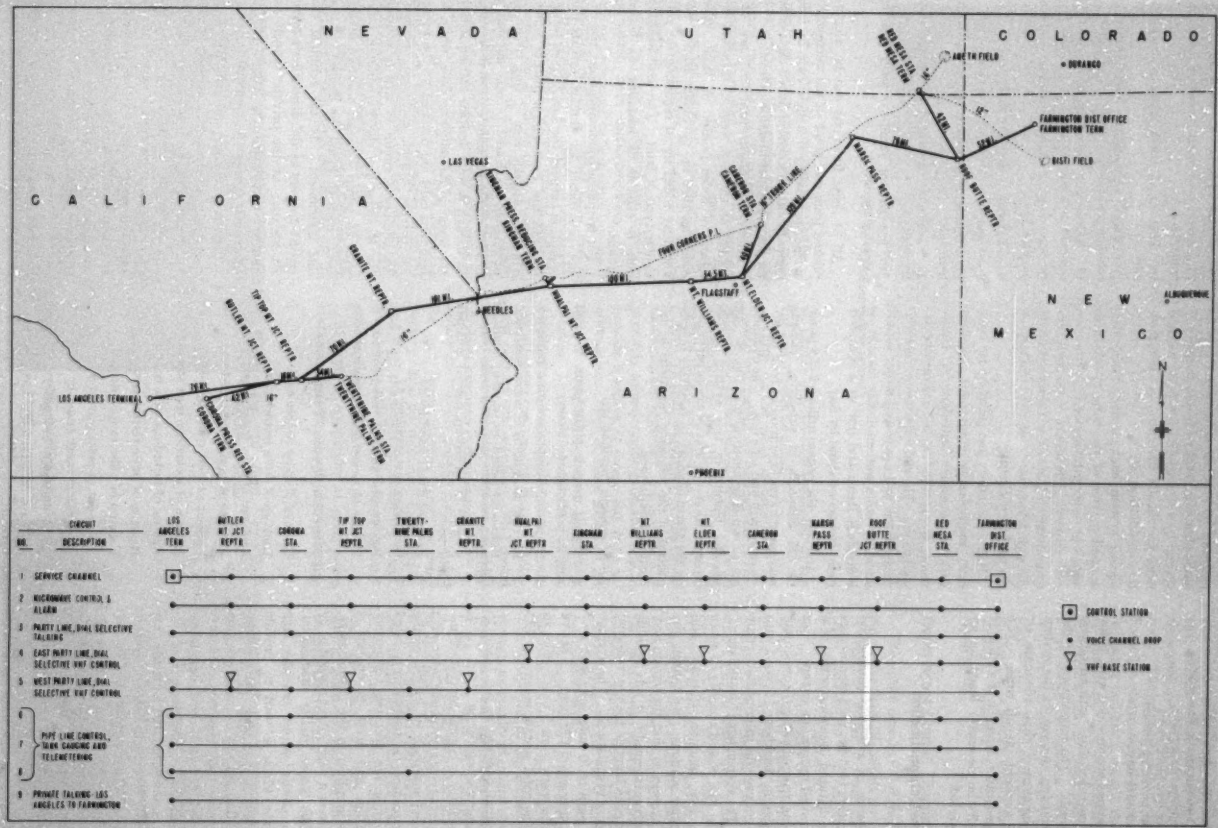


Fig. 8. Microwave communication system route

1. Five open-close flow valves.
2. Three close-automatic flow valves.
3. Incoming line block valve.
4. Outgoing line block valve.
5. Set-point incoming pressure.
6. Set-point outgoing pressure.

These stations are provided with 35-kw stand-by generator units with automatic transfer panels. The importance of the pressure-reducing stations in the system makes it imperative that adequate air supply be available at all times. The stand-by generator capacity is sufficient to carry all the necessary loads for an indefinite period.

Microwave Communication System

The absence of available commercial communication facilities for a greater part of the line dictated the construction of a privately owned microwave system. This system is constructed along the mountain tops, creating a "backbone" system with laterals to connect the intermediate pump stations. The system has 15 installations, as shown in Fig. 8, and consists of seven terminals, five junction stations, and three repeater stations. By routing the backbone system so that the terrain may be used advantageously, it was possible to increase path lengths, thereby reducing the number of installations. The average backbone path is 70 miles long with several paths in excess of 100 miles. The five lateral paths vary from 8 to 42 miles in length. The total system mileage is 795 miles.

The system operates in the 1,850- to 1,990 megacycle operation fixed band. The transmitter is crystal-controlled, using a 90-megacycle crystal oscillator with stability and accuracy of 0.01% of the carrier frequency. Conservatively rated stages provide the necessary frequency multiplication and drive to operate a lighthouse amplifier in the output stage, which delivers a peak power output of 20 watts under pulse carrier modulation. The high power output provides high system gain which is used to provide better fading margins and lower cost installation through the use of lower gain antennas, and a transmission line of higher loss. In using crystal-controlled transmitting equipment, a statement of frequency measurement need be made only once every 6 months in order to comply with Federal Communications Commission requirements. Also, transfer to stand-by equipment is accomplished in minimum time since no "drift-on" time is required for the stand-by unit to become stabilized.

The microwave multiplexing system performs its function of applying 1 to 25 audio channels to the r-f (radio-frequency) system by the time-division (TD) method more specifically described as pulse position modulation (PPM). In a TD multiplexing system, each channel is allocated a small portion of the time in which to accept or reject intelligence to or from the main carrier. The failure of the multiplex equipment on a leg circuit does not affect through transmission of that channel on the backbone circuit.

Full r-f stand-by and stand-by for common multiplexing equipment, together with the necessary fault-detecting equipment, is provided at all microwave installations. The radio-frequency receiver, transmitter, and associated power supply stand-by units automatically transfer as individual items and not as a complete radio-frequency stand-by bay.

Initially, the system utilizes 9 of a possible 25 channels in supplying circuits necessary to the operation of both microwave system and pipe line. The channels are shown diagramatically in Fig. 8. An unusual feature in this system is the incorporation of a microwave supervisory on the alarm channel, to facilitate remote control of the following functions at indicated installations.

All-Microwave Installations

1. Voltage drop test (all power reduced 10% for predetermined times).
2. Exercise of stand-by common multiplex.
3. Lock-up of stand-by common multiplex.
4. Release of locked-up stand-by common multiplex.
5. Release of locked-up normal common multiplex.
6. Exercise of east stand-by r-f.
7. Lock-up of east stand-by r-f.
8. Release of locked-up east stand-by r-f.
9. Release of locked-up east normal r-f.

All Microwave Repeater Installations

1. Exercise of pulse restore east.
2. Lock-up of pulse restore east.
3. Lock-out of pulse restore east.
4. Exercise of west stand-by r-f.
5. Lock-up of west stand-by r-f.
6. Release of locked-up west stand-by r-f.
7. Release of locked-up west normal r-f.
8. Exercise of pulse restore west.
9. Lock-up of pulse restore west.
10. Lock-out of pulse restore west.
11. Exercise of north stand-by r-f.
12. Lock up of north stand-by r-f.
13. Release of locked-up north stand-by r-f.
14. Release of locked-up north normal r-f.
15. Exercise of north stand-by common multiplex.
16. Lock-up of north stand-by common multiplex.
17. Release of locked-up north stand-by common multiplex.
18. Release of locked-up north normal common multiplex.
19. Exercise of stand-by power audible monitor.
20. Exercise of stand-by generator.
21. Lock-up of stand-by generator.
22. Telemeter generator 1 (voltage and frequency).
23. Telemeter generator 2 (voltage and frequency).
24. Energize of stand-by tower beacon (Marsh Pass only.)

All of these functions can be accomplished from any point in the microwave system by dialing a 2-digit code for the desired location followed by a 2-digit code for the particular function desired.

Microwave alarm functions are recorded at the Los Angeles terminal by means of codes printed on tape. The code identifies the particular station followed by coded signal reporting the particular type of fault. The fault-locating unit is capable of handling six coded groups.

For telephone communications, one circuit is utilized for duplex private-line operation between the Los Angeles terminal and the Farmington district office. The Los Angeles end of this circuit is terminated on leased private automatic branch exchange facilities, and it operates as though it were another station within the terminal for outgoing calls, while all incoming calls are handled by the terminal private branch exchange operator. An additional circuit is used to connect the terminal and the Farmington office with all pump stations, selective dialing being used to signal the desired location and instrument. At the terminal this circuit appears at the dispatcher's turret for his use; at the private branch exchange switchboard, for connection to telephones within the terminal.

Two voice channels are used to control very-high-frequency base stations located at all microwave repeater installations. The additional voice channels are used in party line arrangements to provide for the pipe line supervisory control, telemetering, and remote tank-gaging signals. A maximum of 35 narrow-band frequency-shift voice-frequency carriers are transmitted over a single voice channel between 300 and 3,000 cycles per second. Each audio-frequency carrier transmitter

Fig. 9. Outdoor station layout

can be shifted in frequency to either a mark or space frequency. The mark frequency is 12.5 cycles above and the space frequency 12.5 cycles below the channel frequency of 300 cycles per second, while at the 3,000-channel frequency the mark and space frequencies are 25 cycles above and below the channel frequency.

For supervisory control and telemeter signaling, each audio-frequency transmitter is arranged to be in space frequency condition during no-signal input and in mark frequency condition when keyed. The associated receiver reflects simultaneously the condition of its transmitter; the complete absence of an input signal will cause the receiver to remain in mark frequency condition, thereby providing an alarm function where required.

Voice-frequency carriers for supervisory control and telemeter signals between the Los Angeles terminal and the various pipe line distribution points are transmitted over voice channels leased from the telephone company. Business telephones are installed at each of the distribution points to permit maintenance personnel to call the Los Angeles terminal.

Three of the eight mountain-top repeaters are served by commercial power. A propane-fueled stand-by engine with automatic transfer panel is installed at each of these three locations.

With no power available at five of the mountain top repeaters, the operation of privately owned engine-generator sets to provide a source of power was the only alternative. Four of these locations are usually inaccessible during the winter months. The power plants and respective accessories are engineered to operate without attention for 3-month periods. Identical primary and stand-by liquid-cooled diesel engine-generator units are provided at these locations in combination with an automatic transfer panel. Each unit has its own engine-starting panel, starting battery, lube oil regulator, lube supply, and dual primary and secondary filters. A common fuel system with 4 months' capacity is provided.

Microwave equipment at each pump station is housed in a steel-frame asbestos-clad insulated building. At the Farmington district office sheathing and brick were used to conform to the local building code. The repeater-station equipment is housed in standard steel-frame metal-clad insulated buildings, except the Tip Top Mountain repeater equipment, which is housed in a concrete block building having a prefabricated concrete roof to comply with United States Forestry Service requirements. The Los Angeles microwave equipment is housed within the Terminal Building adjacent to the supervisory equipment.

Supervisory Control Equipment

The entire pipe line system, except for the feeder stations, which are automatic, is remotely operated from the terminal office building in Los Angeles County, Calif. Remote operation is effected by means of supervisory control and telemetering equipment installed in the terminal building and at each station. A separate supervisory control system operating over a duplex audio tone channel is provided for each pump station, pressure-reducing station, and valve station. The required channel space for the pump and pressure-reducing stations is provided by the microwave system previously discussed. Leased voice circuits provide channel space for the valve stations in the Los Angeles area.

The supervisory control equipment employed is of the all-relay pulse count type. The equipment in the terminal building consists essentially of groups of telephone-type relays and associated control panels mounted in and on upright cabinets. The cabinets will, with ultimate expansion of the pipe line system, comprise three walls of the dispatcher's office. Generally one cabinet is required for each station, although all the valve stations require only one cabinet and the initial station with its booster pumps and tank manifold requires three cabinets.

Fig. 10 (left). Outdoor motor installation

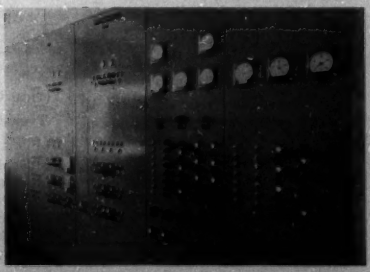


Fig. 11 (right). Local control panels

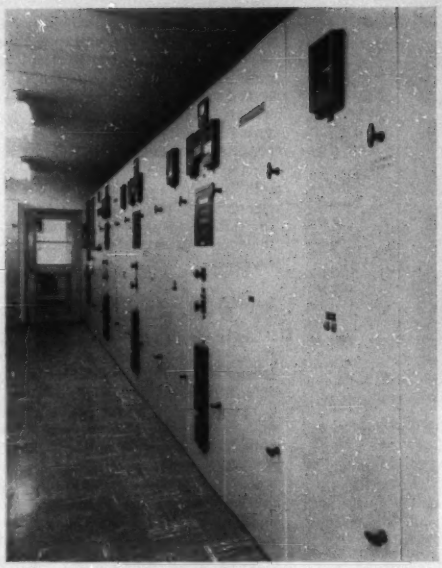


Fig. 12 (left). Switchgear assembly

Fig. 13 (above). Pressure-reducing station

The cabinets are arranged in order, according to the physical layout of the pipe line. The control panels with push buttons, indicating lamps, and mimic piping layouts are mounted on the front of the cabinets. The telemeter receivers are mounted directly above the control panels.

The sealed supervisory relay cases are mounted on the rear of the cabinets on hinged panels. Access space is provided at the rear of the cabinets for testing and maintenance of the equipment.

The supervisory control equipment at a station consists of a set of telephone-type relays and associated-power-type interposing relays to transfer the signals to the power equipment being controlled. The equipment at the pump stations is installed in a switchgear auxiliary cubicle at one end of the switchgear line-up; at the pressure-reducing stations it is installed on the rear panel of the local control cabinets and at the seven valve stations in outdoor weatherproof cabinets.

Each supervisory control system is divided into "points" of supervision and/or control. A separate point is required for each device under supervision or control. A point will usually have three lights on the control panel. One of these is a white light which indicates that the point has been selected, either by the dispatcher or automatically, when an alarm or change is reported. The other two indicate the position of the supervised device; in the case of an alarm function these are green (normal) and red (alarm). For a valve the lights are green (valve closed); amber (valve open); and green and amber (valve in intermediate position). For a motor the lights are green (off) and amber (on).

The points with control as well as supervision are also provided with push buttons or selector push buttons, as required.

The supervisory signals consist of combinations of pulses called codes. To reduce the number of pulses required to identify a particular point, the supervisory control systems with more than ten points are divided into groups with ten points comprising a group. Each group and each point within a group has a particular code consisting of one to ten pulses. The operation and supervision codes consist of three to five pulses and are common for all points.

Typical Control Operation

Typical control operates in the following steps:

1. The dispatcher pushes the individual control push button of the desired function to be performed.
2. The control station transmits the group selection code.
3. The remote station selects the proper group and transmits a group check code identical to the group selection code.
4. Upon receipt of the group check code the control station compares it with the original code sent out and, if identical, transmits the point selection code.
5. The remote station selects the proper point and transmits a point check code.
6. The control station receives the point check code and, if this is identical to the point selection code sent out, it lights the associated white light on the control panel indicating that the point has been selected; it then transmits the operation code.
7. Upon receipt of the operation code the remote station energizes the appropriate power-type interposing relay performing the directed function.
8. When the controlled device changes position, the remote station transmits a supervision code.
9. Upon receipt of the supervision code, the control station changes the indicating lights on the control panel and transmits a reset pulse which resets the equipment at both locations.

If for any reason the selection codes and check codes do not agree, the equipment will automatically reset, and the operator must again depress the push button of the function desired to be performed.

Since the pulses are transmitted at the rate of 13 per second, the entire operation just described takes from 2 to 5 seconds, depending on the particular point selected.

Uncalled for changes in position of supervised devices are automatically re-

Fig. 14. Local control cabinet pressure-reducing stations

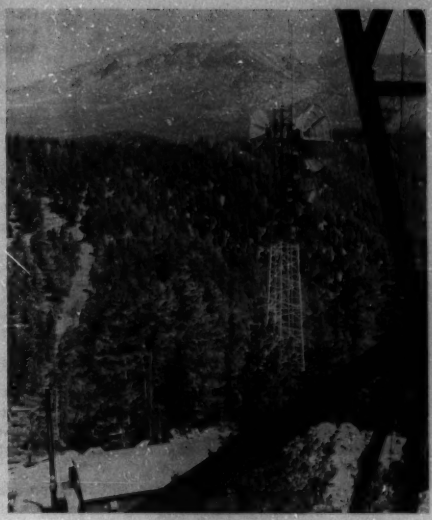


Fig. 15. Mountain-top microwave installation

ported. The operation for this type of function is similar to a control operation, except that the selection codes are originated by the remote station; the check codes are originated by the control station; and the operation code is omitted.

In the case of an alarm-type function automatically reported, the light changes on the control panel are accompanied by the sounding of an alarm. The audible alarm is silenced by depressing the alarm reset button. A change in position of a device of a nonalarm type or a change initiated by the dispatcher is not accompanied by an audible alarm.

A master check button is provided for each system. Operation of a master check button will cause the supervisory control system to check each point in sequence. Each point when checked will momentarily light its corresponding white lamp.

In case of communication failure, each supervisory control system affected reports a communication failure alarm and holds its indications of station-operating conditions. After communications are restored, the supervisory equipment automatically corrects any indications in error During the outage, the stations continue to operate unless shut down by local protective devices.

The supervisory control systems provide the following control and supervision:

All Main Line Units

1. "Start" control and supervision of suction valve.
2. Supervision of discharge valve.
3. Supervision of mechanical trouble.
4. Supervision of electrical trouble.
5. "Stop" control and supervision of unit breaker position and unit lockout.

All Pump Stations

1. Supervision of incoming and outgoing scrapers.
2. Supervision of station valves with control of station block valves.
3. Supervision of station case and discharge pressures.
4. Supervision of station alarms: instrument air pressure, sump level, power failure, and station battery.
5. Control of emergency stop.
6. Raise-lower control of discharge pressure set point.
7. Supervision of remote start cutoff.

Initial Pump Station Only

1. Control and supervision of booster pumps.
2. Control and supervision of tank manifold valves.
3. Control and supervision of tank mixers.

Pressure Reducing Stations

1. Automatic-close control and supervision of three automatic control valves.
2. Open-close control and supervision of five manual control valves.
3. Raise-lower control of incoming and outgoing pressure set points.
4. Control and supervision of relief tank pump.
5. Supervision of incoming and outgoing scrapers.
6. Supervision of station valves with control of station block valves.
7. Supervision of auxiliary generator.
8. Supervision of station alarms: instrument air pressure, sump level, incoming line pressure, outgoing line pressure, and relief tank level.
9. Supervision of remote start cutoff.

Delivery Valve Stations

1. Control and supervision of delivery valves.
2. Supervision of pressure and flow alarms.

An unusual feature is the remote set point control. The set point changer installed at a station consists of a 2-winding reversible motor positioning a precision regulator. The output of the regulator (3 to 15 pounds per square inch air) is fed to a set point bellows located in the discharge pressure controller. The air signal is also fed to a telemeter transmitter.

The gearing is arranged to produce a 1% adjustment for each 1-second operation of the motor. To adjust the set

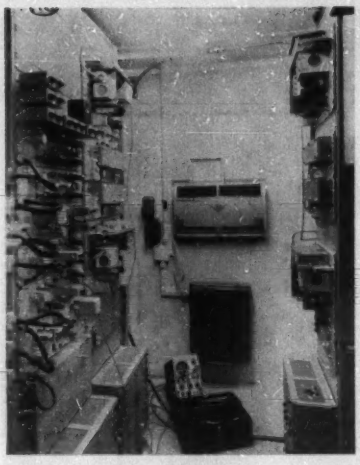


Fig. 16 (left). Communication equipment racks

Fig. 18 (right). Supervisory relay cabinet assembly

point, the dispatcher turns the set point selector push button to the desired position (raise or lower) and pushes it. Operation of the supervisory control equipment results in the motor being energized for a period of 2 seconds, resulting in a 2% change of the set point. Any amount of adjustment desired can be obtained by repeated operation of the raise-lower push button. The set point control scheme is shown in Fig. 7.

Telemetering and Tank Gaging

The telemetering equipment is of the impulse duration type with a 5-second period. Each telemetered quantity is transmitted on a continuous basis over a simplex audio tone channel.

The telemeter transmitters are of explosion-proof design and are grouped outdoors on a common rack. Local telemeter receivers working directly off the transmitters are provided at each station for use when local control is required.

The telemetering receivers at the terminal building are of the miniature type and are installed above the control panels in the supervisory control cabinets. The majority of the receivers are recording, and some are equipped with high and low alarm lights.

The telemetering equipment is self-synchronizing and will automatically correct its readings upon restoration of communications after an outage. During an outage, all receivers affected will read full-scale indicating to the dispatcher their in-operation.

The following quantities are telemetered from an intermediate pump station: station suction pressure, station case pressure, station discharge pressure, flow, gravity, and discharge pressure set point. The initial pump station has, in addition to the foregoing, the flows of the two incoming feeder lines and pressures related to the booster pumps.

Telemetered quantities received from a pressure-reducing station include incoming and outgoing flow, incoming and outgoing pressure, incoming and outgoing pressure set point, and gravity. Also telemetered is the position of the automatic friction tube control valve in operation. Since only one of the three automatic valves is in an intermediate position at any one time, a single telemetering system with automatic selection serves all three. Telemetered quantities from the delivery valve stations include pressure, gravity, and flow from required locations.

A remote reading tank-gaging system is provided for the storage tanks at the initial station and the relief tanks at the two pressure-reducing stations. Each tank is equipped with liquid level gages and averaging thermometers. Two re-

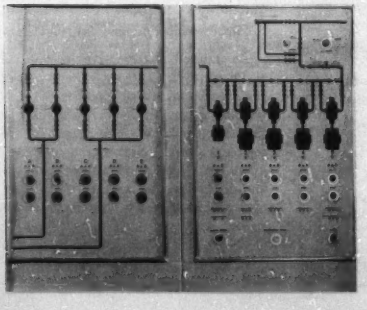


Fig. 17 (left). Supervisory control panels

Fig. 19 (right). Delivery valve station

ceivers are installed, one at the initial station and one at the Los Angeles Terminal Building. The entire system works on a party line basis, and readings from all tanks can be taken at either receiver.

The system employed is of the pulse code type. Transmitted information is digital rather than analog, thus eliminating the possibility of errors being introduced by the transmission system. To take a reading, the dispatcher dials the function and tank desired, by means of a conventional telephone-type dial. After about 5 seconds, a reading appears on a visual read-out panel giving the tank number and liquid level or temperature. Liquid level high and low alarms are automatically reported.

Figs. 9 through 19 show various indoor and outdoor views of the system.

Summary

The Four Corners pipe line is designed for completely unattended operation. The main line pump stations, pressure-reducing stations, delivery valve stations, booster pumps, and tank-filling and suction valves are designed for remote operation. Most of the feeder stations are equipped for automatic custody transfer operation.

The design of the communications and control features incorporated into the system provides for safe and reliable operation of the pipe line from the dispatcher's office in the terminal building located at the western terminus of the line. The design provides the controlling point with all the necessary telemetered quantities, pressures, flows, gravity, and set points necessary for the remote operation of the facilities. All necessary local protective devices are provided at all locations to permit safe operation during communications or supervisory control equipment failures.

Appendix. Device Designations

Designations are based on the proposed AIEE Standard No. 68.[7]

Station Functions

Station functions are given in the following, with the device numbers.

1—main breaker control
1—control switch at breaker
1A—indoor control switch
1B—outdoor control switch
1CSW—automatic operation cutoff switch
5—main breaker tripping relay
8B—main breaker tripping power disconnect switch
8C—main breaker closing power circuit breaker
8SF—station functions disconnect switch
8STC—shunt trip coil unit valve feeder breaker
20-1A—friction tube automatic control valve No. 1-A
20-5M—friction tube manual control valve No. 5M
20D—station discharge valve
20IS—incoming scraper valve
20ISC—incoming scraper valve closing contactor
20ISM—incoming scraper valve motor
20ISO—incoming scraper valve opening contactor
20KI—outgoing scraper No. 1 kicker valve
20K2—outgoing scraper No. 2 kicker valve
20L—incoming scraper lead-in valve
20OS—outgoing scraper valve
20S—station suction valve
20S-1A—solenoid valve control valve No. 1A
20S-5M—solenoid valve control valve No. 5M
27DC—battery undervoltage relay
30—Annunciator drop
33SA-1—arriving scraper detector switch No. 1
33SA-2—arriving scraper detector switch No. 2
33SD-1—departing scraper detector switch No. 1
33SD-2—departing scraper detector switch No. 2
42CAI—air compressor starter
42PSV—sump pump starter
42YL—yard light contactor
45HA—combustible gas alarm switch
47—undervoltage or reverse phase voltage relay
51—overcurrent relay
52—main circuit breaker
63AI—instrument air pressure switch
63AIM—air compressor control pressure switch
63D—station discharge pressure switch
63IL—incoming line pressure
63OL—outgoing line pressure
63RVI—incoming relief valve pressure switch
63RVO—outgoing relief valve pressure switch
69—remote control cutoff switch
71RTL—low-level relief tank switch
71RTH—high-level relief tank switch
71SU—sump level switch
71SUM—sump pump control level switch
74R—alarm reset relay
80RV—flow to relief tank switch
86—station lockout relay
88CAI—air compressor motor
88PSU—sump pump motor
90MR—set point control motor-running winding
90MS—set point control motor-starting winding
96—auxiliary generator-running contact
EM—telemeter receiver electromagnet
GM—telemeter receiver gear motor
M—telemeter transmitter cam motor
MS—telemeter transmitter keying switch

Unit Functions

The following device numbers are for unit 2. Device numbers for the other units are the same except that the prefix number is the same as the unit number.

201—start-stop control push button
204—sequence-starting relay
205—sequence-stopping relay
206—starting circuit breaker
208—control power disconnect switch
219—start period timing relay
220D—discharge valve
220PVT—pump vent valve
220S—suction valve
226M—motor-winding temperature relay
226PC—pump case temperature switch
227—a-c undervoltage relay
230—annunciator drop
238MI—motor inboard bearing temperature switch
238M2—motor outboard bearing temperature switch
238P1—pump inboard bearing temperature switch
238P2—pump outboard bearing temperature switch
239PVB—pump vibration detector switch
239SF—pump seal leak switch
242—running breaker
246—phase balance relay
248—over-all sequence check timing relay
249—thermal overload relay
249IT—thermal overload relay instantaneous trip contact
251—overcurrent relay
251IT—overcurrent relay instantaneous trip contact
263PC—pump case pressure switch
263S—pump suction pressure switch
269—permissive switch
273—motor space heater relay
286—lockout relay
286XE—electrical trouble relay, lockout
286XM—mechanical trouble relay, lockout
294E—electrical trouble relay, nonlockout
294M—mechanical trouble relay, nonlockout

Supervisory control interposing relay device numbers are the same as the numbers of devices they control except when prefixed with SC.

References

1. Four Unattended Pipeline Pumping Stations. **R. S. Cannon, T. V. Bockman.** *The Oil and Gas Journal*, Tulsa, Okla., Dec. 7, 1953, pp. 178–79.

2. Electricity's Contribution to Modern Pipelining. **M. A. Hyde.** *The Westinghouse Engineer*, Pittsburgh, Pa., Sept. 1953, pp. 154–61.

3. Remote Operation of Pipe-Line Pumping Stations. **W. A. Darr, M. A. Hyde.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, Sept. 1954, pp. 190–98.

4. Microwaves for Pipelines, Part VIII, Supervisory Control and Telemetering. **C. B. Lester.** *The Pipe Line News*, Bayonne, N. J., July 1953, pp. A1–A8.

5. Basic Circuitry for Electrically Powered Pipe-Line Pump Stations Under Automatic or Remote Control. **M. A. Hyde, W. A. Darr.** *AIEE Transactions*, pt. II, vol. 74, Mar. 1955, pp. 4–14.

6. Remote Controlled Unattended Pump Stations Utilized by Great Lakes Pipe Line Company. **P. R. Madden, A. P. Riahi.** *PEPA Paper*, Petroleum Electric Power Association, Dallas, Tex., May 16, 1956, pp. 51–59.

7. Proposed Device Numbers and Functions for Pipeline Pump Stations Under Automatic or Remote Control. *AIEE Standard No. 68*, Feb. 1958.

The Life Expectancy of Class A Random-Wound Motor Insulation As Determined by AIEE Standard No. 510 Test Procedure

AIEE COMMITTEE REPORT

TEST PROCEDURES for determining the relative life expectancies of different insulation systems for random-wound machines, using either complete motors or models, called motorettes, are provided for by AIEE no. *510*.[1] The procedure calls for subjecting the motorettes to successive periods of high temperature in ovens, followed by vibration and exposure to moisture, and an overvoltage check. The cycle is repeated until the test unit fails under voltage. The tests are conducted at three or more temperatures, and are made on a sufficient number of samples to obtain the desired statistical accuracy. The tests on motors are similar, but the heating, vibration, and voltage exposures are all obtained at the same time by operating the motors on a frequent starting or reversing cycle. The results of the tests are given in the form of a life-temperature chart, as in Fig. 1.

To make this procedure fully useful, and a basis for standards, it is necessary to establish the life expectancy that AIEE no. *510* determines for widely used insulation system that has proved satisfactory in extended service use. Then those who are responsible for standards, after they have decided what the relative insulation life required for the standards should be as compared with the familiar system, will be able to set a numerical value for the hours of life under the no. *510* procedure that an insulation system should have to satisfy the requirements.

With this purpose in view, the committee responsible for the development of AIEE no. *510* undertook to carry out a series of tests to establish by this procedure the life expectancy of the present-day widely used class *A* random-wound motor insulation system, that is giving highly satisfactory service in motors built under American Standards. This present system has evolved as the result of many years of experience. Formerly cotton-covered wires and paper slot liners were used. About 1937, film-covered wires began to replace the cotton. Another marked change occurred about 1952 when the polyester Mylar* came into use for slot liners. The slot liners used in the present-day systems contain a layer of polyester sheet (synthetic) material, which has markedly superior life under extreme moisture conditions, as compared with the cellulosic materials (paper or varnished-cloth) in the slot liners of the class *A* insulation system employed up to about 5 years ago. In Europe, where the motor insulation systems are similar to ours, this superior life of the synthetic materials under extreme conditions very recently has been recognized by designating the all-synthetic system as class *E* for which the limiting temperature has been set at 115 C (degrees centigrade). In the United States, this superior system is still designated as class *A*, with a limiting temperature of 105 C. Thus, present-day motors built under the American Standards are more conservatively rated than European motors. This means that, other things being equal, the life expectancy of standard American motors at their rated temperature rise, corresponding to the usual 1.15 service factor (50 C), should be about twice as long as that of the comparable European motors at their rated-load temperature rises (60 or 65 C).

There are a good many variations in the details of the present-day materials and processes used by different manufacturers. However, the differences are minor, and probably they do not cause variations any greater than those resulting from other causes inherent in the test procedure.

Test Program

Nine laboratories participated in the program, which was conducted in the following manner.

The insulating materials to be used in the tests were supplied to all the laboratories from a common stock, made up of the following items:

Copper Wire. 0.040-inch no. 18 Awg (American Wire Gauge) heavy Formvar covered wire, with a total build of 3 mils. Using this material, coils of 14 turns of two wires wound in hand were used in the motorettes.

Slot Liners. 0.005-inch Mylar laminated to 0.005-inch rag paper. The Mylar was on the inside of the liners. The length of overhang at each end of the slot, while not specified, was approximately 1/4 inch.

Varnish. Oil-modified phenolic, 2 dips and bakes. The viscosity, and the time and temperature of bake were not specified, but typical values are 400 cycles per second at 25 C, with a bake of 3 hours at 150 C.

Wedges. Vulcanized curved fiber 1/32-inch thick.

Tying Cord. No. 12 cotton cord wax treated.

Sleeving. Oil-varnished-glass sleeving.

Phase Separators. 0.015-inch rag insulating paper.

Motor Leads. Silicone rubber was used to facilitate handling. The leads do not form a part of the system under test.

All these items are typical of those regularly used in American general-purpose class *A* insulated motors, but they are not identical with those used by any one manufacturer, so far as known. The one material that varies most widely between different manufacturers probably is the insulating varnish.

Each laboratory assembled these materials into motorettes, or into motors, in accordance with their own practices, following the general rules specified in AIEE no. *510*. All parts of the motors and motorettes not listed were obtained independently.

Each laboratory performed the sequence of tests prescribed in AIEE no.

Paper 58-1334, recommended by the AIEE Rotating Machinery Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 27, 1958; made available for printing December 8, 1958.

Members of the AIEE Working Group on Insulation for the Rotating Machinery Committee are: P. L. Alger, *chairman*; R. L. Balke, D. R. Blake, H. P. Boettcher, R. L. Brancato, J. P. Dexter, T. J. Gair, G. P. Gibson, W. T. Gordon, L. Greer, J. L. Kuehlthau, L. P. Mahon, G. L. Moses, G. A. Mullen, W. W. Pendleton, W. B. Penn. Messrs. Blake, Boettcher, Brancato, Dexter, Gair, Gibson, Mahon, Mullen, and Pendleton participated in the program.

The committee wishes to express appreciation for the great amount of time and effort put into this test program by the nine committee members and their laboratory associates who participated in the test program. Also, special acknowledgment is due to Mr. L. P. Mahon and his associates in the Canadian General Electric Company, who prepared a program for the International Business Machines Corporation *650* computer, and used this to make all the statistical calculations required in evaluating the test results. This report of the committee's work was prepared by the chairman and secretary, but it includes many suggestions made by committee members after seeing a preliminary draft.

* Registered trade-mark E. I. Du Pont de Nemours & Company, Inc.

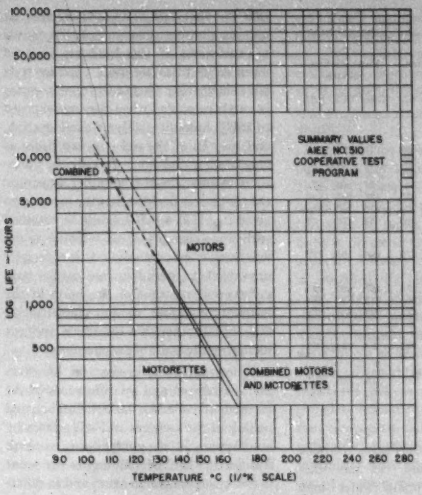


Fig. 1. Combined life-temperature curves as determined by AIEE no. 510 co-operative test program

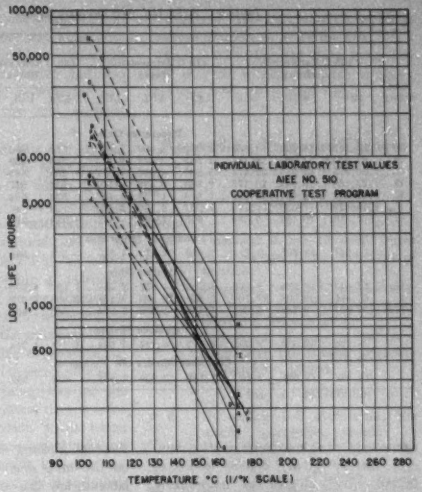


Fig. 2. Individual life-temperature curves as determined by AIEE no. 510 co-operative test program

510 at three or four different temperatures (approximately 130, 150, and 170 C), and used 10 motorettes or 5 motors to determine each point. Six of the laboratories tested motorettes only, two tested motors only, and one tested both motors and motorettes, making ten sets of data in all. Another laboratory, *I*, planned to make both kinds of tests but unforeseen circumstances prevented the motorette tests from being completed, so the limited data obtained are not included here.

Table I. Motorette Tests

Laboratory	Life, Hours 105 C	Life, Hours 170 C	Type of Failure	Amp Fuse Size	Test Voltage Between Wires
A*	13,742	191	Wire	Neon	60
B†	21,016	140	Wire	Neon	120
C	30,715	220	Wire	Neon	120
D	15,188	187	Wire	1/8	50
E‡	9,606	235	Wire and ground	1	120
G	7,730	61	Wire and ground	5	120
F	14,122	235	Wire and ground	Neon	120

* Laboratory A used a low oil-modified phenolic varnish that was not the same as those used by the other laboratories.

† All the laboratories except B used phase separators in the form of grids cut from sheet, whereas B used separate strips in each slot.

‡ Laboratory E used 17 instead of 14 turns in the coils, giving a materially tighter fit in the slots and somewhat greater length of wire. Also, the varnish used by this laboratory, obtained from another source, had a lower viscosity than normal (about 140 cycles per second at 23 C).

Test Results

The over-all results of the program are summarized in Table I, and in Fig. 2. The principal conclusions reached are:

1. The average life of usual class *A* random-wound motor insulation systems, when tested in accordance with the existing AIEE no. *510* procedure, and extrapolated to 105 C, is roughly 12,000 hours, including both motors and motorettes, but values found by different laboratories varied from about 5,000–60,000 hours.

2. The average life expectancy is halved for each 11 C increase in temperature, approximately, over the range between 105 C and 170 C.

3. The motors gave nearly the same average life as the motorettes at 105 C and about 200% at 170 C, with about 12 C for half-life of the motors, and about 10 C for half-life of the motorettes.

Discussion of the Results

Table II gives the results obtained in the ten independent sets of tests, and Fig. 2 shows these in chart form. Inspection of the data reveals a rather wide variation in the results obtained by the different laboratories. This is not surprising when it is considered that a change of only 5 C in the temperature corresponds to about 40% change in the life. After reviewing all the information, the committee decided that plausible reasons can be assigned for the major variations in results, as indicated in the following summary of the differences in the procedural details followed by the various laboratories.

As noted in Table I, two of the laboratories used a lower test voltage between wires than the 120 volts used by the other five laboratories. This came about because the committee decided to change the requirement of 50 volts in the present version of AIEE no. *510* to 120 volts, in order to obtain more definite end points, particularly for fiber-covered wires. This change was made too late for all the laboratories to follow.

In the next edition of AIEE no. *510*, 120 volts will be called for in this test.

It appears from Tables I and III that differences in the fuse size were a major cause for the differences in life obtained by the different laboratories. There is a considerable voltage drop in a neon lamp when only a small leakage current flows, so that the applied voltage falls, and the unit has time to dry out before the initially small leakage current that flows during a voltage check test can reach a destructive value. Also, when a small fuse is employed, the circuit is opened before any damage occurs, and a later test shows no indication of failure, so that the unit could be continued on the test program in the belief that the indication of failure was false.

When a larger fuse is used, the leakage current has a high enough value, and con-

Table II. AIEE No. 510 Summary Table

Laboratory	105 C Extrapolation, Hours			130 C Extrapolation, Hours			150 C Extrapolation, Hours			170 C Extrapolation, Hours		
	Life	UCL	LCL	Life	UCL	LCL	Life	UCL	LCL	Life	UCL	LCL
Motorettes Only												
A	13,742	19,225	9,807	2,253	2,689	1,901	619	650	588	191	202	180
B	21,016	33,564	13,159	2,535	3,042	2,112	556	559	553	140	167	117
C	30,715	41,985	22,470	3,805	4,426	3,271	855	886	825	220	235	205
D	15,188	18,944	12,177	2,365	2,588	2,160	626	627	624	187	204	171
E	6,606	8,021	5,441	1,614	1,775	1,467	589	604	575	235	245	226
F	14,122	17,650	11,269	2,501	2,790	2,243	726	746	706	235	247	225
G	7,730	10,153	5,884	997	1,088	912	230	241	220	61	72	52
Combined Motorettes												
	11,543	13,707	9,721	2,014	2,178	1,863	578	584	572	186	195	177
Motors Only												
H	12,446	22,575	6,862	3,162	4,077	2,453	1,187	1,200	1,176	487	602	394
I	63,886	109,998	37,104	9,550	12,589	7,258	2,458	2,673	2,261	715	782	653
J	5,007	8,656	2,896	1,359	1,839	1,004	535	608	471	229	236	222
Combined Motors												
	17,114	30,520	9,507	3,474	4,656	2,592	1,111	1,214	1,016	394	434	357
Combined Motors and Motorettes												
	11,485	13,891	9,495	2,150	2,349	1,968	649	659	639	218	230	207

tinues long enough, to cause positive damage, thereby giving definite proof of failure. Also, failures are more likely to occur because of creepage over the ends of the slot tubes when very small fuses are used. It appears from the test data that the average life may be only one half or one third as great when the 5 ampere fuses are used, as compared with that when Neon lamps are employed. Consideration of all factors leads to the belief that no material change in life will occur in motorette testing, if the size of fuse is increased above 5 amperes. From the data in this paper alone, it is known that a smaller fuse than 5 amperes may be adequate, but it appears that when fiber-covered wires are tested, as in higher temperature systems, a somewhat larger fuse will be needed than is required for enameled-wire systems. It is expected that further evidence on this subject will be brought out in AIEE papers now being planned to cover additional test data.

On this basis, it was decided to include in the next revision of AIEE no. *510* the specific requirement that 1-ampere fuses be used in motorette testing for film-covered wires, and that 5-ampere fuses be used when the wire is fiber-covered.

Variations in the degree of humidity and the ways in which the samples were exposed are important also. In laboratory *B*, the motorettes were placed vertically in the humidity chamber, while in the other laboratories they were horizontal. In laboratory *E*, they were upside down, so that they were protected from drip. In laboratories *D* and *F* all metal parts of the humidity chamber and motorettes were either made of stainless steel or thoroughly plated, so that water dripping on the units was pure. In the other laboratories, the dripping water may have carried with it a certain amount of rust. Laboratory *E* mounted ten motorettes on 2 plates of Mycalex and placed them in the humidity chamber in a horizontal position, unit side down. This mounting technique permitted more abundant condensation on the units, and its localization in drops, then the procedure followed by other laboratories. Laboratory *E* removed the motorettes from the humidity chamber and placed them in plastic containers to limit evaporation, before applying the electrical tests. The test voltage was applied to all ten motorettes at the same time. In Laboratory *C*, electrical tests were made on the motorettes in air, after they were brought out from the humidity chamber, whereas in the other laboratories except *E* the voltage was applied with the unit still in the humidity chamber.

In all cases, there was water in the bottom of the humidity chamber that was vaporized by immersion heaters, so that extensive condensation occurred in drops on the test units, but there was no way to be sure that the amount of condensation was the same in all laboratories. It appears that the differences in life caused by variations of this kind are reduced when larger fuses are used. Earlier tests have shown that the test life of motorettes is roughly doubled when they are exposed to 100% humidity without condensation, and may be of the order of five times as long when tested dry. Despite the variations inherent in the humidity exposure, the committee decided that the procedures given in AIEE no. *510* should be retained with some changes in the wording to define them more precisely. It is especially important to maintain the water temperature above that of the units in the humidity chamber, so that condensation will actually occur on the units, and not merely on the walls of the chamber.

Other variations in the test life were undoubtedly caused by differences in the methods of assembly, varnish dipping, and baking of the samples and still others by inaccuracies in the temperature control. A difference of 2 or 3 C may easily occur in different parts of an oven, and in different parts of a motor winding, and these alone would account for differences of 15–25% in insulation life. Errors of this type should be averaged out by testing a sufficient number of samples, however. Measurements showed that the amplitude of motion of units differently placed in the vibration test varied between 4 and 12 mils more or less, as compared with the 8 mils that are specified. Here again the errors should be averaged out in testing a considerable number of units.

On the whole, it appears that the degree of variation can be reduced appreciably in the future by more accurately defining the procedures of AIEE no. *510*, especially the fuse size as mentioned previously. It also appears that the variations that occurred in the present test program were generally on the high side, so that tests made in the future in accordance with the proposed revision of AIEE no. *510* should give somewhat shorter life on motorettes.

In the motor tests, the windings are subjected to transient overvoltages, when the motor is switched on and off, and these

Table III. Motor Tests

Laboratory	Life, Hours 105 C	Life, Hours 170 C	Type of Failures	Amp Fuse Size	Slot Liners	Type of Motor Phase	Hp	Poles	Volts
I	63,886	715	Between phases	6½	5/32-inch cuff	1	1/6	4	110
H	12,446	487	Wire	35	3/8-inch straight	3	10	6	440
J	5,007	229	Wire and between phases	40	5/16-inch cuff	3	5	4	440

Laboratory I used no phase separators. Laboratory H used separate strips in each slot, and Laboratory J used phase separators in the form of grids.

overvoltages are concentrated on the terminal coils. The motorettes do not experience any similar transient voltage, but they are subjected to 120 volts between turns throughout the entire winding. These differences, together with the differences between the stresses caused by starting and reversing the motors, as compared with those produced by vibration of the motorettes, give ample cause to expect a somewhat different life to be obtained for motors than for motorettes. Further differences are caused by the different procedures for the single-phase and 3-phase motors, and by the choice of the size, number of poles, and voltage rating of the motors tested. A smaller number of motors are used to check each point than is the case for motorettes.

These factors help to explain why there is a wider dispersion in the results of the motor tests than for the motorettes. Laboratory *H*, in making some of their tests on single-phase motors, applied a repeated surge impulse voltage of 600 volts, as a periodic check on one group of motors, in addition to the normal tests. This procedure did not change the observed life appreciably. For this and other reasons, the committee decided not to make repeated surge impulse tests a part of the normal procedure in AIEE no. *510*. It is thought that, with the normal type of industrial contactor used in switching, the transient voltages that occur in a random manner during switching are adequate, and give fairly comparable stresses to those produced by the turn to turn test on the motorettes.

However, it is recognized that, when tests are made on higher temperature and higher voltage insulation systems, the use of moisture as a means of searching out insulation defects caused by aging may not be as effective as in the low voltage class *A* tests just described. For this reason, the use of repeated surge impulse tests on an optional basis is permitted by AIEE no. *510*.

It is expected that additional tests will be made as time goes on to evaluate higher voltage and higher temperature random-wound motor insulation systems, following the procedures of AIEE no. *510*, and that experience so obtained may indicate further desirable changes in the procedure. It is believed that AIEE no. *510* after the revisions indicated in the foregoing and explained more completely in the appendix to this paper have been made, will be quite satisfactory as a means of evaluating the merits of different random wound motor insulation systems. In general, it is believed that the use of motorettes is particularly desirable as a means of evaluating new materials and complete systems for general-purpose use, while tests on motors will be desirable to check and extend the motorette results, and also to evaluate insulation systems for particular kinds of service and environments.

It should be noted that the twisted pair aging test specified in AIEE no. *57*[3] gives a typical life of about 30,000 hours at 105 C for the same wire used in this series of motor and motorette tests. Such a difference is natural, and to be expected, because in AIEE no. *57* failure is obtained by applying a relatively high voltage, in air, in the absence of condensation, instead of by a low voltage with condensation on the units as in AIEE no. *510*. It is desirable to prescribe the system test conditions in such a way as to make the life expectancy less for the complete insulation system than it is for the component materials when tested separately, because the system tests are inherently more expensive, and also subjected to many more causes of variation, so that they often need to be repeated. It is the opinion of the committee that on the broad grounds of economy and practicality it is desirable to adjust the requirements of AIEE no. *510* to give a usual life expectancy under test conditions of about 10,000 hours, more or less, when extrapolated to the limiting temperature of normal operation.

Conclusions

The committee believes that the overall results of tests, as given in Table I and Fig. 2, show that it is entirely feasible to base the requirements for random-wound motor insulations in future American motor standards on test results obtained by following the procedures of AIEE no. *510*. So far as the test program reported in this paper goes, it appears that the present-day insulation systems for general-purpose random-wound motors built in accordance with American Standards, have an extrapolated life of somewhere between 5,000 and 15,000 hours at 105 C, when tested in accordance with the procedures of AIEE no. *510* as they now are. However, the committee does not consider the data reported here sufficiently consistent to warrant standards being based upon these figures. Accordingly, the committee has decided to undertake a new test program using motorettes, to be carried out with exactly the same materials as in the present program, but following the modified procedures prescribed in the proposed revision of AIEE no. *510*. The appendix of this paper gives the text of the revisions that are proposed in AIEE no. *510*.

It is hoped that the data from this new program will be available in about one year from the present time.

Appendix. Proposed Changes in AIEE No. 510

Methods of Evaluation

Pages

6 Paragraph 3, line 10 should read: *A*, *B*, *F*, and *H* for each type of specimen.

6 Paragraph 3, line 17 should read: in which any new insulation, (omitting, material or).

6 The second sentence of the paragraph with the asterisk preceding it should be deleted.

Part I. Motorettes

6 In section 1–2 Motorettes, the second paragraph should be changed to read:
Specifically it is recommended that for the purpose of testing random wound motor insulation, two standard sizes of motorettes are available. Experience indicates that comparable results are obtained whichever size is used. Since the smaller one is economical, it is suggested that it be used when new test programs are undertaken.

8 Add the following sentence to the last paragraph:
It is suggested that the slot liners be made 1/4 inch longer than the slots in order to prevent failure by creepage over the ends.

Temperature and Exposure Time

11 In section 2–4 Moisture Exposure, in the first paragraph insert these two new sentences after the first sentence:
The condensation should consist of fine drops over the entire surface of the unit. To secure this, it is desirable to mount the units in a horizontal position.

11 Under note, delete the words room temperature and add to the sentence the following: temperature of the units themselves, so that condensation will occur on them and not merely on the walls of the chamber.

Voltage Checks

11 In section 3–1 the voltage between conductors should be changed from 50 to 120.

12 Add the following sentences to the second paragraph, face up.
To obtain definite and uniform indication of failure, a fuse shall be placed in series with each unit. The size of the fuse shall be 1 ampere if the wires being tested are film covered, and 5 amperes if the wires are fiber covered. The use of neon lamps in place of fuses is not permissible.

Part II. Motors

15 Section 2–5 Voltage Checks is to be completely deleted and this new 2.5 inserted in its place.

2.5 Voltage Checks. The test motors are to be run during the heat exposure periods at their highest rated nameplate voltage. The end point of the motor life in these tests is fixed by its electrical failure under the applied voltage, as indicated by the melting of a fuse that is connected in series with each motor. The current rating of the fuse should be 5 amperes, or at any rate not less than 1 ampere and not more than 10 amperes. The use of neon lamps or similar indicating devices in series with the motor, in place of or in addition to fuses, is not permissible, because even a small current through such a device creates a considerable voltage drop, thereby lowering the voltage that is impressed on the motor, and prolonging the time before failure occurs.

It is important that the motors should be restarted on the heat exposure phase of each cycle immediately after humidification, and the voltage check tests, described in the following paragraph, be made with a minimum of elapsed time.

Since the insulation of the individual wires in a random winding is not adequately checked by the application of normal line frequency voltage, it is desirable that some type of surge voltage be applied to the windings also. Normally, this is provided by the transient overvoltage surge that occurs each time voltage is applied to the motor by closing the line contactor. The magnitude of the surge will be greater when the contactor is reclosed immediately after being opened, as in the operation of motor reversal. Also, the magnitude will vary materially, depending on the type of contactor, the circuit capacitance as measured at the motor terminals, and the time phase angle of the voltage at the instant the contactor is closed. Therefore, the amount of overvoltage applied will vary in a random manner, on repeated starts or reversals.

It is recommended that the magnitude of switching voltage surge be measured, and steps be taken, if necessary, to adjust the value to approximately 600 volts crest to ground for motors of 110 voltage rating or less, and to 2,000 volts for motors rated 440 volts. It is not important, nor is it practical, to control this overvoltage accurately. It is expected that proper values will always be obtained if normal industrial type magnetic contactors are used, with a time delay in reclosing of 2–5 cycles.

As an alternative to reliance on the overvoltage due to switching—or in addition thereto—a repeated surge impulse voltage may be applied to each motor, immediately following each period of humidity exposure, with visible condensation on the winding.

For this purpose, the surge voltage should be applied for a period of one minute to each phase of the motor winding in turn. The magnitude of the voltage surge should be 600 volts crest to ground for motors rated 110 volts, and 2,000 volts crest to ground for motors rated 440 volts, with intermediate values varying linearly in accordance with the equation:

$$\text{Crest voltage} = 150 + 4\ (\text{Rated Voltage}) \quad \text{approximately}$$

If the motor does not fail under this check test, it should be immediately returned to the operating and heat-aging cycle, as indicated in the foregoing.

References

1. Test Procedure for Evaluation of Systems of Insulating Materials for Random-Wound Electric Machinery. *AIEE Standard no. 510.* Nov. 1956.

2. Test Procedure for Evaluation of the Thermal Stability of Enamelled Wire in Air. *Ibid.*, Jan. 1959.

A Relay-Type Feedback Control System Design for Random Inputs

A. M. HOPKIN
MEMBER AIEE

P. K. C. WANG
NONMEMBER AIEE

EARLIER literature has suggested the design of the so-called optimum relay control system, where the term "optimum" is based on the system having the fastest possible response in terms of reducing the error to minimum following a step or ramp input.[1–4] It has been pointed out that these systems are optimum only in this special sense, and that their response to other types of input may be undesirable.[5]

Attempts have been made to combine predicting devices with the concept of the optimum relay control system to produce systems which possess good response for a large class of inputs.[6–10] The limitation on these schemes is that they become difficult to realize in their given form for systems of an order higher than second, with the result that the concepts have not been carefully explored.

This paper is the report of an attempt to accomplish some further exploration of the possibilities of optimum predictor relay control systems by exploiting a slightly different viewpoint, and to set up a simple model of the system on an analog computer so as to obtain some idea of the behavior of a system in the presence of random inputs.

Basic Assumptions

A relay-type control system can be designed to operate with random input signals if the following assumptions are satisfied:

1. The structure of the input signal and associated noise is such that the input signal can be predicted for a moderate distance into future time in terms of a power series whose term coefficients are functions of the present value of the total input and a finite number of its time derivatives. For the simple noise-free case the prediction function might be in the form of a truncated Taylor's series whose variable would be distance from the present into future time, and whose term coefficients would be present values of the input signal and a finite number of its derivatives.

2. For relay forcing the controlled process response and its $(n-1)$ order derivatives can be adequately described by powers series approximations in the time domain.

3. The criterion for satisfactory system behavior is as follows: The system shall operate to reduce predicted error and its $(n-1)$ order derivatives to some small nominal value in a minimum time. The symbol n is the order of the differential equation describing the controlled process behavior.

4. For certain types of controlled process, with oscillatory modes, the magnitudes of error and its various order derivatives are so bounded that only one mode of relay logic is necessary for control. This restriction is placed in this paper only to simplify discussion; it can be removed at the expense of complicating the decision logic required.

The Form of the System

The general form of system operation is shown in Fig. 1, and is as follows: The command signal $v(t)$ is assumed to be made up of two components, a desired signal $r(t)$ and an undesirable noise signal $n(t)$. The total signal $v(t)$ is fed into a composite filter F, which produces signals representing the coefficients of

Paper 59-218, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing December 8, 1958.

A. M. Hopkin and P. K. C. Wang are with the University of California, Berkeley, Calif.

This report is based on work done as a result of financial support supplied by National Science Foundation Grant *G-1701*.

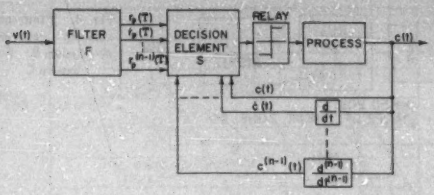


Fig. 1 (left). Simplified control system block diagram

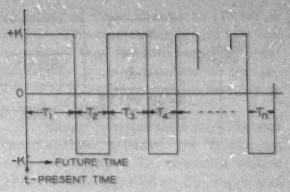


Fig. 2 (right). Relay-forcing function

the predicted desired input power series

$$r_p(t,\tau)=A(t)+B(t)\,\tau+C(t)\,\tau^2+\dots \quad (1)$$

where $r_p(t,\tau)$ represents a prediction made at the present time t of the value of r at some future time $(t+\tau)$, with τ representing the prediction time.

With the use of the series form of the predicted values for the desired system response to maximum forcing in either direction, the predicted switching intervals $\tau_1, \tau_2, \tau_3 \dots$ of a relay forcing function of the form shown in Fig. 2 can be computed continuously. Again, the reference for $\tau_1, \tau_2, \tau_3 \dots$ is always present time t. The system response for future time $c_p(t,\tau)$, may be formulated as a Taylor's series, using $c(t)$, $\dot{c}(t)$, $\ddot{c}(t)$, etc., as the initial condition values determining the series coefficients. The criterion for determining the switching times is based on the fact that at the end of the nth switching interval, the predicted error $e_p(t,\tau_1+\tau_2+\dots\tau_n)$, and its $(n-1)$ order derivatives should equal zero. For the cases where the resulting equations lead to multiple choices of τ_1, τ_2, τ_n, the selection must be made to make all of the switching intervals real and positive, and to bring the error to zero in minimum time.

The error conditions result in a family of equations of the form:

$$\begin{aligned}
&e_p(t+\tau_1+\tau_2+\dots+\tau_n)\\
&\quad=r_p(t+\tau_1+\tau_2+\dots\tau_n)-\\
&\qquad c(t+\tau_1+\tau_2+\dots\tau_n)=0\\
&\dot{e}_p(t+\tau_1+\tau_2+\dots+\tau_n)\\
&\quad=\dot{r}_p(t+\tau_1+\tau_2+\dots\tau_n)-\\
&\qquad \dot{c}(t+\tau_1+\tau_2+\dots\tau_n)=0\\
&e_p^{(n-1)}(t+\tau_1+\tau_2+\dots+\tau_n)\\
&\quad=r_p^{(n-1)}(t+\tau_1+\tau_2+\dots\tau_n)-\\
&\qquad c^{(n-1)}(t+\tau_1+\tau_2+\dots\tau_n)=0 \quad (2)
\end{aligned}$$

If assumption 4 is valid there will be n switchings and n equations required to describe a system whose equations of motion are of nth order.

The time τ_1, which represents the time of the next relay reversal, is of primary interest. When τ_1 has a value greater than zero, the present switching mode is correct for the moment. When τ_1 equals zero, reversal of the relay is called for, and if τ_1 has only values less than zero, the system is in error, and reversal of the relay is called for. The case where τ_1 has no real solution represents a momentary form of input which the system cannot follow because of its inherent limits. A careful study is necessary to determine the proper switching mode to make the best of this bad situation.

In general a computer can be set up with a decision logic element S to operate on present input and output information in such a way as to strive to bring the predicted future error to some nominal small value in minimum time.

Discussion of a Particular System

A better idea of the considerations involved in designing and operating this type of system may be obtained by considering the particular system shown in Fig. 3, consisting of a linear process such as a simple motor with the Laplace domain transfer function

$$G(s)=\frac{1}{s(Ts+1)}$$

The future time response of the process for a fixed amplitude step forcing of $\pm K$ applied at the present time t is

$$c_p(t,\tau)=c(t)\pm K\tau+T(\mp K+\dot{c}(t))(1-e^{-\tau/T}) \quad (3)$$

The predicted response at time $t+\tau_1+\tau_2$ of the same system to a fixed amplitude step-forcing function of $\pm K$ applied at the present time t followed by a step reversal of forcing function to $\mp K$ at time $t+\tau_1$ is given by

$$\begin{aligned}
c_p(t,\tau_1,\tau_2)=c(t)+T\dot{c}(t)\pm\\
[K(\tau_1-\tau_2)+KT-2KTe^{-\tau_2/T}]+\\
[\pm KT-T\dot{c}(t)]e^{-(\tau_1+\tau_2)/T} \quad (4)
\end{aligned}$$

The predicted input for time $t+\tau_1+\tau_2$ is given by

$$r_p(t,\tau,\tau_2)=r(t)+\dot{r}(t)[\tau_1+\tau_2]+\frac{\ddot{r}(t)}{2}(\tau_1^2+2\tau_1\tau_2+\tau_2^2) \quad (5)$$

So the predicted error at $t+\tau_1+\tau_2$ is given by

$$\begin{aligned}
e_p(t,\tau_1,\tau_2)=e(t)+\dot{r}(t)(\tau_1+\tau_2)+\\
\frac{\ddot{r}(t)}{2}[\tau_1^2+2\tau_1\tau_2+\tau_2^2]-T\dot{c}(t)\mp\\
[K(\tau_1-\tau_2)+KT-2KTe^{-\tau_2/T}]+\\
[\mp KT+T\dot{c}(t)]e^{-(\tau_1+\tau_2)/T} \quad (6)
\end{aligned}$$

To find the initial values of $e(t)$, $\dot{e}(t)$, $\dot{r}(t)$, and $\ddot{r}(t)$ which must exist to bring the error and error rate to zero simultaneously in time τ_2 after switching, expressions are written for the predicted error and error rate at the switching time $(\tau_1=0)$ and set equal to zero.

$$\begin{aligned}
e_p(t,0,\tau_2)=e(t)+[\dot{r}(t)\pm K]\tau_2-\\
[\pm TK+T\dot{c}(t)][1-e^{-\tau_2/T}]+\\
\frac{\ddot{r}(t)}{2}\tau_2^2=0 \quad (7)
\end{aligned}$$

$$\begin{aligned}
\dot{e}(t,0,\tau_2)=\frac{\partial e_p(t,0,\tau_2)}{\partial\tau_2}\\
=\dot{r}(t)\pm K-[\pm K+\dot{c}(t)]e^{-\tau_2/T}+\\
\ddot{r}(t)\tau_2=0 \quad (8)
\end{aligned}$$

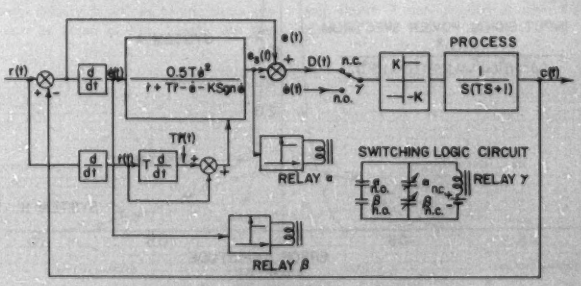


Fig. 3. Block diagram for a particular system

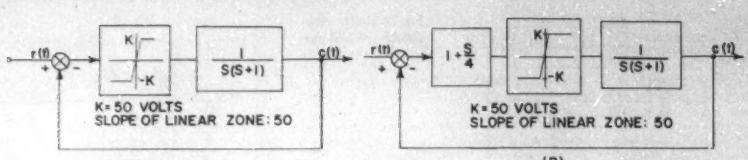


Fig. 4. Motor-positioning systems. A—System B. B—System C

Equations 7 and 8 should then be solved simultaneously to give the desired relationship between $e(t)$, $\dot{r}(t)$, $\ddot{r}(t)$, and $\dot{e}(t)$.

However, because the equations are transcendental, their solution is difficult. This difficulty is removed by making use of assumption 2, and replacing the exponential terms in equations 7 and 8 by the first few terms of their equivalent power series to obtain

$$e_p(t, 0, \tau_2) = e(t) + \dot{e}(t)\tau_2 - \left[\frac{\pm K - \dot{e}(t)}{2T}\right]\tau_2^2 + \frac{\ddot{r}(t)}{2}\tau_2^2 = 0 \quad (9)$$

$$\dot{e}_p(t, 0, \tau_2) = \dot{e}(t) - \left[\frac{\pm K - \dot{e}(t)}{T}\right]\tau_2 + \dot{r}(t)\tau_2 = 0 \quad (10)$$

Equations 9 and 10 can be solved simultaneously to give the switching boundary relationship

$$e_s(t) = \frac{0.5T\dot{e}^2(t)}{[\mp K + \dot{e}(t) + T\ddot{r}(t)]} \quad (11)$$

Equation 11 sets up the boundary relation between the present values of $e(t)$, $\dot{e}(t)$, $\ddot{e}(t)$ and $\ddot{r}(t)$ at which reversal of the forcing function should occur. To reduce the number of quantities involved, $\ddot{e}$ can be replaced by $\dot{r}(t) - \dot{e}(t)$ to give expression

$$e_s(t) = \frac{0.5\dot{e}^2(t)}{[\dot{r}(t) + T\dot{r}(t) - \dot{e}(t) \mp K]} \quad (12)$$

Since the sign of the forcing term $\pm K$ is determined by the additional requirements that τ_2 be positive and of minimum duration, it is shown in the Appendix that the sign of $\pm K$ should be chosen to be the negative of the sign of the error rate. Assigning $(-\text{signum } \dot{e})$ as the sign of K, yields

$$e_s(t) = \frac{0.5T\dot{e}^2(t)}{\dot{r}(t) + T\ddot{r}(t) - \dot{e}(t) - K \text{ Signum } \dot{e}(t)} \quad (13)$$

Note that $e_s(t)$ is negative if $([\dot{r}(t) + T\ddot{r}(t)])$ is zero and $\dot{e}$ is positive. However, in general the value of $e_s(t)$ for a given value of $e(t)$ will vary as $[\dot{r}(t) + T\ddot{r}(t)]$ varies.

As shown in the Appendix by use of phase-plane concepts, if $\dot{e}(t)$ and $e(t)$ have the opposite sign and $e(t)$ is greater than $e_s(t)$, then the sign of the forcing $\mp K$, should be positive, and if $e(t)$ is less than $e_s(t)$, then the sign of the forcing should be negative. This suggests the rule: *Apply the forcing function having the sign of* $D(t) = e(t) - e_s(t)$.

The phase plane plots shown in the Appendix demonstrate that tangential intersections of $r(t)$ and $c(t)$ without error overshoot cannot occur for initial conditions $e(t)$ and $\dot{e}(t)$ have the same sign. For these initial conditions the sign of the forcing should be the same sign as the error rate. The switching logic can be formalized as follows:

Condition 1: If $e_s(t)$ and $\dot{e}(t)$ have the same signs, give the system forcing function the sign of $\dot{e}(t)$.

Condition 2: If $e_s(t)$ and $\dot{e}(t)$ have opposite signs, give the system-forcing function the sign of $D(t)$, where

$$D(t) = e(t) - e_s(t) \quad (14)$$

Experimental Results

A motor-positioning system of the type described was set up on the analog computer in order to obtain experimental data concerning its response to various inputs. This system was called system A, and is shown in the block diagram of Fig. 3.

Two other systems were set up for comparison with system A. One, designated as system B, consisted of a motor positioning system driven by a high-gain amplifier saturating at the levels $\pm K$; see Fig. 4(A). The second,

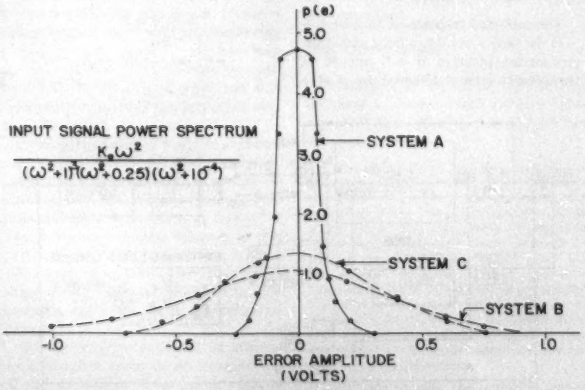


Fig. 5. Comparison of error amplitude probability density curves for systems A, B, and C

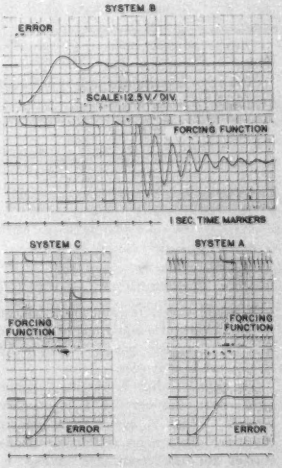


Fig. 6. Step responses of systems A, B, and C

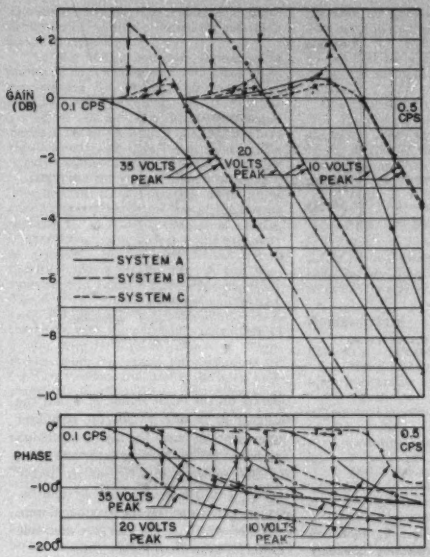


Fig. 7. Frequency response of systems A, B, and C

system *C*, was the same as system *B* except that the sharply saturating amplifier was preceded by a derivative-compensating element so adjusted as to give essentially the same step response as system *A*; see Fig. 4(B).

Each system was tested by supplying its input with a signal generated by passing the output of a Gaussian random signal generator through a low-pass filter. This type of random test signal was selected because it was considered fairly representative, and because it was conveniently generated.

The resulting input and output data were recorded and the error amplitude probability density functions were computed from these data; they are plotted in Fig. 5.

In interpreting the error amplitude probability density function curves as a measure of the "goodness" of the system response it must be recognized that the ideal error probability density function would be a unit impulse at the error amplitude of zero, and that, in general, assuming a symmetrical distribution of error about the origin, the control system which produces the narrowest error amplitude probability density function curve for a given random input has the best response. Comparison of the error amplitude density functions of the three systems shows that for the random test inputs the predictor-controlled relay system has a far better measure of behavior than the other two systems.

It is interesting to compare the amplitude density functions of the three systems as given in Fig. 5 with the step responses of these systems as plotted in Fig. 6. This comparison would lead one to suspect that the step response is not necessarily a good measure of the behavior of a nonlinear systems response to a random input.

Frequency response tests were also run for the three systems; the results are shown in Fig. 7. The predictor system behavior was degraded by the response of the prediction coefficient filter at frequencies above those expected in the random input signal.

In general it was felt that test results on the predictor system were good when the predictor was able to predict with reasonable accuracy for a length of time representing the maximum expected value of time τ_3 required to bring the error to zero. This maximum expected time depends upon the nature of the statistical structure of the input signal. When the nature of the input was changed so as to render the predictor less effective, the degradation of system performance was immediately noticed. Fig. 8 shows predicted responses to give some idea of the effectiveness of the prediction scheme used for different Gaussian inputs. For Fig. 8(A)

$$\frac{K\omega^2}{(\omega^2+1)^2(\omega^2+10^{-4})(\omega^2+0.25)}$$

and for Fig. 8(B)

$$\frac{K\omega^2}{(\omega+1)^4(\omega^2+10^{-4})}$$

Other Types of Systems

A similar procedure may be used to derive the switching criterion for the case of a relay-driven two-time-constant system with a linear Laplace transfer function of the form

$$G(s)=\frac{1}{(T_1s+1)(T_2s+1)} \tag{15}$$

The general solution for the system output at time $(t+\tau)$ resulting from a step input of $\pm K$ applied at time t is given by

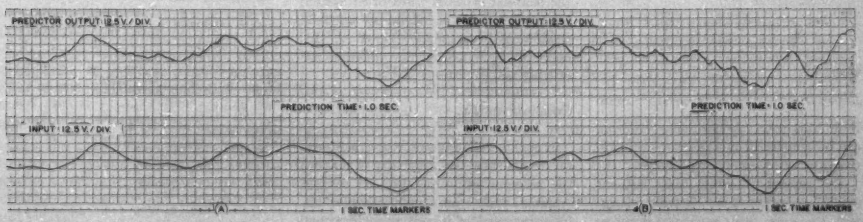


Fig. 8. Response of Taylor series predictor to Gaussian input signal with power spectral density

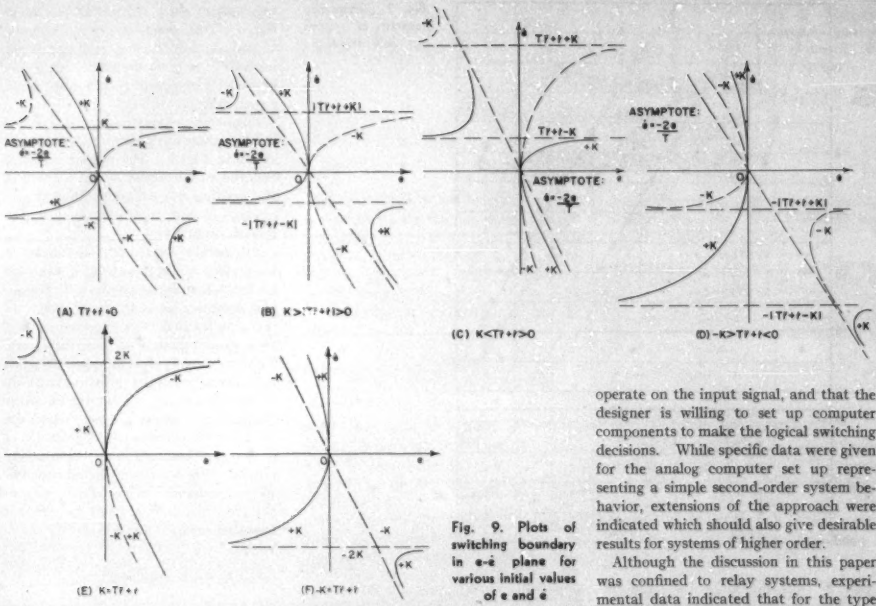


Fig. 9. Plots of switching boundary in e-ė plane for various initial values of e and ė

$$c(t,\tau)=\frac{T_1T_2}{T_1-T_2}\left[\left(\frac{c(t)}{T_2}+\dot{c}(t)\right)e^{-\tau/T_1}-\left(\frac{c(t)}{T_1}+\dot{c}(t)\right)e^{-\tau/T_2}\right]\pm K\left[1+\frac{1}{T_2-T_1}\left(T_1e^{-\tau/T_1}-T_2e^{-\tau/T_2}\right)\right] \tag{16}$$

The procedure used in the preceding case can then be followed with appropriate choice of $\pm$ sign specification to produce positive, real values for τ_2 to give the following switching boundary expression:

$$e_s(t)=-\frac{K\ \text{signum}\ \dot{e}(t)+[T_1+T_2]\dot{e}(t)-I(t)}{2}+1/2\ \text{signum}\ \dot{e}(t)\times\sqrt{[K\ \text{signum}\ \dot{e}(t)+(T_1+T_2)\dot{e}(t)-I(t)]^2-2T_1T_2\ddot{e}^2(t)} \tag{17}$$

where

$$I(t)=T_1T_2\dot{r}(t)+(T_1+T_2)\ddot{r}(t)+\dddot{r}(t) \tag{18}$$

From this point the switching criterion can be established in a fashion similar to that given in the preceding example.

The general procedure holds also for oscillatory second-order systems. However, the power series expansion is not so accurate, particularly for systems with appreciable damping. The considerations discussed in assumption 4 apply here.

The procedure can be extended to third-order systems and higher. In general the criterion must require that the error and its $(n-1)$ order derivatives be zeroed simultaneously for an nth order system. This usually leads to the assumption of $(n-1)$ reversals of the forcing relay. Writing the n equations implied by this statement, setting the time to the first switching τ_1 to zero in each, and eliminating the remaining $\tau_2, \tau_3 \ldots \tau_n$ variables, will give an expression for the switching criteria.

The procedure is straightforward but it leads to complicated algebra. Only the simplest types of third order system have criteria which can be set up on a small analog computer. However, this approach may have promise when the use of a digital computer is justified.

Conclusions

It was felt that the general experimental results show the possibility of making the relay or saturating type of control systems which are quite effective in following random inputs, providing that an adequate predicter can be set up to operate on the input signal, and that the designer is willing to set up computer components to make the logical switching decisions. While specific data were given for the analog computer set up representing a simple second-order system behavior, extensions of the approach were indicated which should also give desirable results for systems of higher order.

Although the discussion in this paper was confined to relay systems, experimental data indicated that for the type of random inputs used, systems with narrow-proportional-band saturating amplifiers operated with the amplifiers in the saturated mode most of the time, and so the mode of operation would be essentially the same if the amplifier were replaced by an ideal relay. The philosophy of control system design outlined in this paper should be of particular interest when other considerations dictate the necessity of using a digital computer in conjunction with the system.

Appendix. Phase Plane Description of Switching Criteria

Since equation 13 represents the condition where instantaneous reversal of the forcing function will ultimately bring the system error and error rate to zero simultaneously, it is the expression for the critical switching boundary in the error-error-rate plane; see Fig. 9. It should be noted that changing the $>$ sign to a $<$ sign in the conditions given in Figs. 9(B) and (D) turns the plane over with respect to the e-axis.

Trajectories and boundaries in this plane are sometimes confusing because the system is not autonomous, i.e., the system has a time-varying forcing component in the form of $r(\tau)$. There is also some difficulty in that the mathematical solution allows the physically impossible case where the time

τ_3 is less than zero. The problem is to select a physically realistic criterion for switching based on equation 13.

The plot of this boundary expression as shown in Fig. 9 is labeled with signs of K, representing the direction of the forcing function after switching, i.e., during the period τ_3. In all cases the solution boundaries in the first and third quadrants (where sign $\dot{e}$ equals sign $\dot{e}$) represent solutions for negative values of τ_3, so there is no realizable optimum switching mode in these quadrants. The best mode of operation under these conditions is to make the sign of the forcing function K the same as the sign of e and $\dot{e}$, so as to reduce error and error-rate as rapidly as possible. In the second and fourth quadrants there are, in general, two solutions for the Taylor's series approximation. The one which leads to simultaneous zero error and error rate in the minimum time is the curve with the maximum magnitude of $\dot{e}$ component for a given value of e. Thus, in the upper half-plane, only solutions for $(+K)$ are meaningful, and in the lower half-plane only solutions for $(-K)$ are meaningful.

Since in the upper half-plane all system trajectories move from left to right, with increasing time, a $(+K)$ switching boundary divides the upper half-plane into two regions with τ_1 and $(-K)$ to the left, and τ_2 and $(+K)$ to the right. Similarly, since all phase trajectories in the lower half-plane go from right to left, the $(-K)$ switching boundary divides it into two regions with τ_2 and $(-K)$ to the left and τ_1 and $+K$ to the right. Thus, in all cases the region to the right of the appropriate boundary calls for $(+K)$, and the region to the left calls for $(-K)$.

The switching criteria just given are based on the truncated series form of expressions for $r_p(t+\tau_1+\tau_2)$, and $\dot{e}_p(t+\tau_1+\tau_2)$ and are valid only when $(\tau_1+\tau_2)$ is small compared to the time constant. This means that for cases where e_p and $\dot{e}_p$ are large, the switching logic may not be effective.

One may select the appropriate value of $\pm K$ for computing the e_s boundaries by the rule: *Make the sign of K in equation 13 the negative of the sign of $\dot{e}$.* The sign of the actual applied forcing function can be determined by the following rules:

1. If sign $e_s(t)$ = sign $\dot{e}(t)$, make the sign of the forcing function the same as the sign of $\dot{e}(t)$.

2. If sign $e_s(t) \neq$ sign $\dot{e}(t)$, make the sign of the forcing function the sign of $D(t)$, where $D(t) = e(t) - e_s(t)$.

Nomenclature

$r(t)$, $\dot{r}(t)$ . . . = desired component of input signal and its time derivatives

$n(t)$ = undesirable noise component of input signal

$v(t)$ = actual input signal $v(t) = r(t) + n(t)$

t = present time

τ, τ_1, τ_2 = future time

$r_p(t, \tau)$ = predicted future desired signal at $(t+\tau)$, predicted at time t

$A(t)$, $B(t)$, $C(t)$ = time-varying coefficients

$c_p(t, \tau)$ = predicted future system response at $(t+\tau)$ predicted at time t

$c(t)$, $\dot{c}(t)$, $\ddot{c}(t)$ = present system response and its time derivatives

$e_p(t, \tau)$, $\dot{e}_p(t, \tau)$ = predicted system error $r_p(t, \tau) - c_p(t, \tau)$, and its time derivative

S = decision logic element

T, T_1, T_2 = system time constant

$e_s(t)$ = time-varying switching boundary value of $e(t)$

signum (x) = sign of x

$D(t)$ = time-varying switching function $D(t) = e(t) - e_s(t)$

References

1. A Phase-Plane Approach to the Compensation of Saturating Servomechanism, Arthur M. Hopkin. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 631–39.

2. Nonlinear Techniques for Improving Servo Performance, D. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 400–21.

3. An Investigation of the Switching Criteria for Higher Order Contactor Servomechanisms, J. Bogert, L. F. Kazda. *AIEE Transactions*, vol. 73, pt. II, July 1954, pp. 118–27.

4. Differential Equations with a Discontinuous Forcing Term, D. W. Boshaw. *Report no. 469*, "Experimental Towing Tank," Stevens Institute of Technology, Hoboken, N. J., Jan. 1953.

5. A Differential-Analyzer Study of Certain Nonlinearly Damped Servomechanisms, R. R. Caldwell, V. C. Rideout. *AIEE Transactions*, vol. 72, pt. II, July 1953, pp. 165–70.

6. A Predictor Servomechanism, N. C. Walker. *M. S. Thesis*, University of California, Berkeley, Calif., 1953.

7. A Taylor's Series Approach to the Design of a Predictor Servomechanism, G. F. Aroyan. *M. S. Thesis*, University of California, Jan. 1957.

8. Feedback Control Systems (book), O. J. M. Smith. McGraw-Hill Book Company, Inc., New York, N. Y., 1958, pp. 505–603.

9. Optimum Design of Final Value Control Systems, R. G. Booton. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Microwave Research Institute, Polytechnic Institute of Brooklyn, New York, N. Y., vol. XI, 1956.

10. Final Value Controller Synthesis, M. V. Mathews, C. W. Stegg. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-2, Feb. 1957, pp. 6–16.

Detector Car History and New Developments

H. W. KEEVIL
NONMEMBER AIEE

THE rail detector car is an assemblage of specialized test equipment mounted on self-propelled railroad cars and is used for the detection of all types of defects that develop in railroad rails. The story of detector cars should start a few years before there were any such cars. Before 1926, the railroads were faced with a serious and fast-growing menace to their very life: the breaking of rails. The source of this menace was the transverse fissure, a very thin crack which would start within the rail and develop outwardly until it caused the rail to break.

Careful rail failure records showed that if a rail developed one transverse fissure the probabilities were that there would be one or more additional fissures in the same rail. Thus if one fissure broke it would create additional strain on the rest of the rail which would then cause the rail to break at the other fissures. This, of course, caused a very dangerous condition because parts of the rail might fall completely out of track. These fissures starting within the rail and developing outward gave no sign of their presence until the rail actually broke.

It is interesting to note that through joint research by the AAR (Association of American Railroads) and the rail manufacturers the cause of the transverse fissure was determined. It was found to be a shatter crack which is formed in the steel as it cools after rolling. By a new controlled cooling process put into service in 1938, these shatter cracks have been eliminated and rail since that date has been practically free of transverse fissures.

Fig. 1 shows large, well-developed transverse fissures. This is the type of defect that was instrumental in the development of detector cars. There are other types of defects that develop in rail which can be detected and removed from track before the rails fail.

Fig. 2(A) shows various sized vertical split heads. This defect is a longitudinal crack that may be a few inches to a full

Paper 59-267, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE–American Society of Mechanical Engineers Railroad Conference, Chicago, Ill., April 7–9, 1959. Manuscript submitted June 5, 1958; made available for printing February 9, 1959.

H. W. Keevil is with the Association of American Railroads, Chicago, Ill.

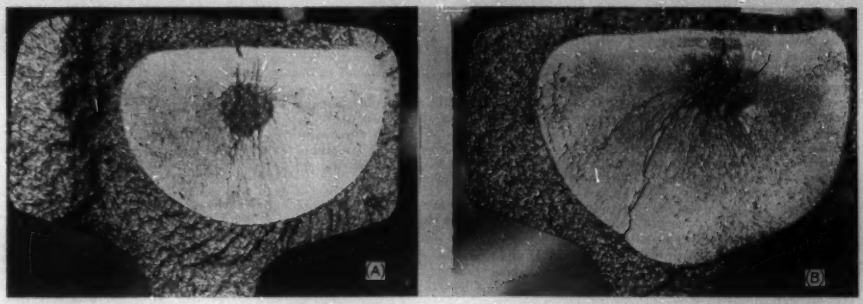


Fig. 1. Large, well-developed transverse fissures

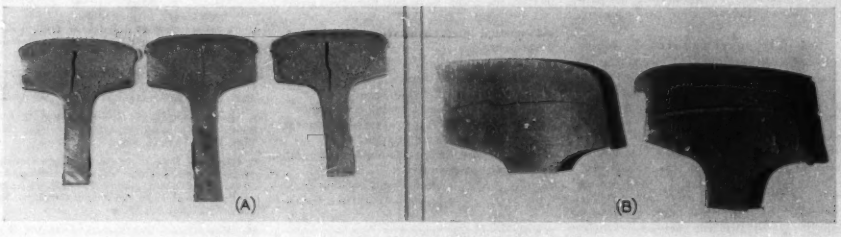


Fig. 2. A—Vertical split heads. B—Horizontal split heads

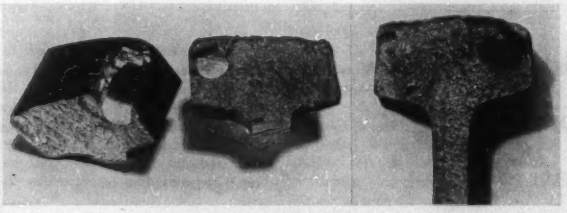


Fig. 3. Example of shell fractures, which usually start at edge of rail

Fig. 5. Butt weld failure developing in gas pressure butt weld: this progressive fracture caused by small inclusion

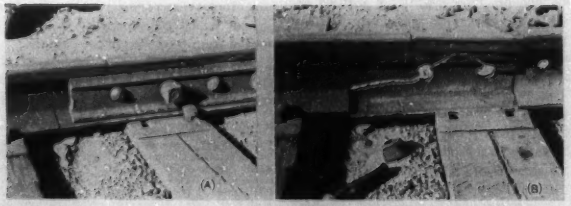


Fig. 4. A—Bolt-hole breaks, which can progress in all directions. B—Same joint, angle bars removed

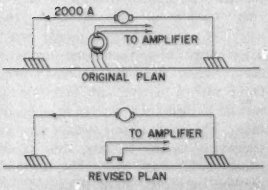


Fig. 6. Electro-inductive system: original plan for drop-of-potential detection

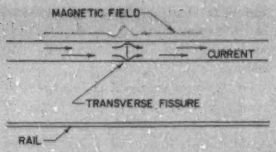


Fig. 7. Magnetic field produced by rail current

rail in length. The crack causes the side of the head to break off.

Fig. 2(B) shows a similar defect known as a horizontal split head because the crack is a horizontal plane. This defect may also progress along the rail for any length. The great danger here is that the defect may turn either upward or downward and progress until it breaks the rail.

Fig. 3 shows progressive fractures which usually start in a little shelling at the edge of the rail. This type of defect is on the increase and has replaced the transverse fissure as the most serious type of defect.

Fig. 4(A) is another serious defect which is usually found in older rail. This crack starts in a bolt hole and can progress in all directions. Fig. 4(B) shows the same joint after the angle bars were removed. The danger here is that a passing train may kick the top piece partly out of the joint, thereby causing a natural derail.

Fig. 5 shows the latest type of defect which develops in a gas pressure butt weld. The small inclusion was the cause of this progressive fracture.

As may be seen, there was real need for a rail-flaw detector. In 1926, W. C. Barnes, engineer of tests on the Rail Committee of the AAR, was delegated to investigate any and all possible means of detecting these defects before the rails failed.

A promising means of detecting internal rail fissures was suggested by Dr. Elmer A. Sperry, of gyroscope fame. He proposed passing a current through the rail longitudinally and measuring the drop of potential between two fixed points. This system was tried in the laboratory and proved successful. For use on a car he proposed using three contacts on the rail which would thus give two equal voltages as long as the contacts were on good rail. These two voltages were opposed so as to produce a balanced input to an amplifier. This automatically balanced out the effect of any fluctuation of the current in the rail but if a set of the contacts bridged a fissure the two current-resistance drops were unbalanced, and thus there was an impulse fed to the amplifier and recorded.

Mr. Barnes recommended that the AAR finance the development and the building of the first detector car by Dr. Sperry. The recommendation was approved and the car was scheduled to be completed by December 1927. The car construction was proceeding well by November 1927, and it was moved to a New York Central Railroad shop for completion and testing.

When car operation was attempted, it was found that there were many difficulties to overcome. One of the major difficulties was the inability to obtain and maintain perfect contact on the top of the rail. It had been thought that the top surface of the rail would be bright, clean, and easy to contact. This proved to be just the opposite of what actually existed. The rail surface, while it may look bright and clean, actually has, in most cases, a hard film of dirt and oxide rolled on it.

Many months were spent trying to devise some means of maintaining good contact on the rail. All of this work was

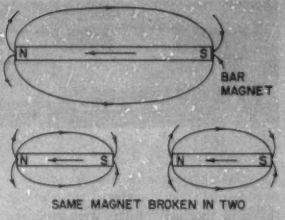


Fig. 9. Basic magnetic principle: magnetized rail showing defect sets up local poles, detectable above rail

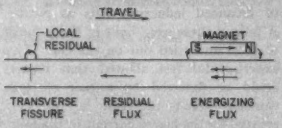


Fig. 10. Principle of residual magnetic detection: progressively magnetized rail produces detectable local fields at points of defect

Fig. 11. First residual magnetic detector car, 1936

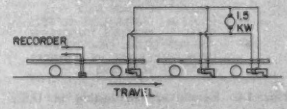


Fig. 12. Principle of first successful residual magnetic detector car

Fig. 8. First electro-inductive detector car

Fig. 13. One of 14 successful residual magnetic detector cars, all still in use

of no avail because as soon as pressures were made great enough to cut through this film, thermal voltages were produced that far exceeded the potentials that were to be picked off the rail. Finally the drop-of-potential method was given up and it was decided to try a pair of balanced induction coils as pickup. A pair of these were built in the summer of 1928 and tried out on the car. This experimental test showed that fissures could be detected by a pair of coils working above the rail in the field created by the heavy current in the rail.

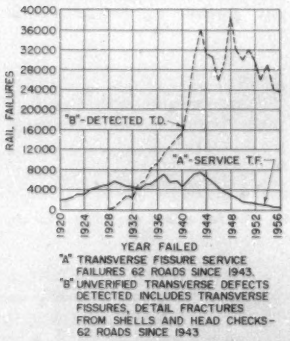


Fig. 14. Results of rail testing in U.S.: Annual service rail failures due to transverse fissures and detected transverse defects as reported by all railroads

Fig. 6 shows the original plan for the drop-of-potential detection, and the induction system that finally proved successful. This change of pickups changed the whole method of detection. Instead of picking a drop of potential off the rail, the pickup coils generated voltages when they passed through any distortion in the magnetic field that surrounded the rail.

Fig. 7 is a crude representation of the magnetic field strength above the rail containing a transverse fissure. This distortion of the magnetic field is due to the change in the direction of the heavy current as it passes around the defect.

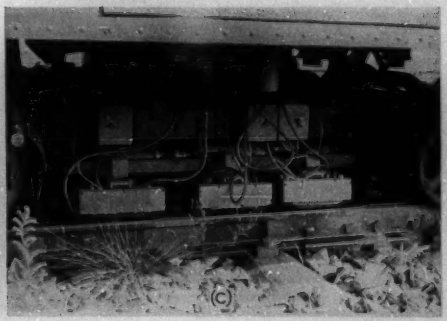
It took several months to build the new induction pickup, and revise the amplifiers and associated equipment. In November 1928, one year late, the Association car was ready for road testing.

Fig. 8(A) shows the AAR car approximately a year after it had originally started working. Many changes and improvements had to be added to make it operate successfully. Fig. 8(B) shows the main brush carriage, set between wheels, which contacted the rail and also carried the induction pickups.

This car, while successful, operated under many difficulties, so Mr. Barnes and the writer were engaged by the AAR to develop an improved flaw-detector car that did not conflict with patents in the art owned by others. During experimentation, it was discovered that a transverse fissure could be magnetized with a small permanent magnet, and a residual field left after the magnet was removed.

It has been demonstrated in physics that a bar magnet, when broken in two, produces two magnets. In the case of the rail, it was magnetized and if there was a defect present, local poles

Fig. 15. A—German ultrasonic detector car. B—Cabin and monitors. C—Crystal carriage and housings



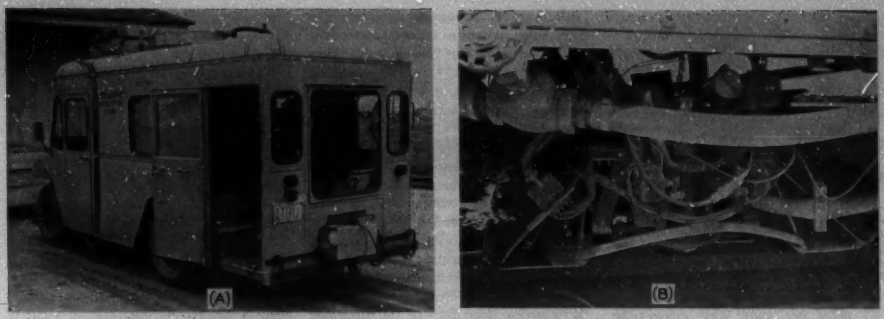


Fig. 16. A—ATSF ultrasonic detector car, operable on track or on highway. B—Detail: one vertical probe used for in-joint detection

Fig. 17. AAR residual magnetic detector car, recently rebuilt with modern test equipment

were set up which were detectable above the rail; see Fig. 9.

Fig. 10 illustrates the principle of residual magnetic detection. The rail is progressively magnetized and a strong residual flux is left within it. At points of defect small local fields are produced outside the rail. The local field can be detected and recorded. This discovery was checked and demonstrated to various railroads. The building of the first residual magnetic detector car was then authorized.

Fig. 11 shows the first residual magnetic detector car, which was put on the road in 1936, and promptly failed to detect transverse fissures. It was determined that a single U-shaped magnet, even though very strong, would not leave a fissure magnetized. It was also determined that a single magnet of any shape would not magnetize a small transverse fissure. From these experiments it was learned that three magnets were needed and that they should be L-shaped with a high rear pole. This was to obviate any strong vertical flux at the leaving end of the magnet. To get three magnets over each rail, it was necessary to go to two cars instead of the original one. This was done by borrowing the tow car from the induction-type car. With two cars it was possible to put three L-shaped magnets over each rail. This combination was the first successful magnetic detector car.

Fig 12 shows the principle of the first successful residual magnetic detector car; and Fig. 13, the actual cars used for this development. This car was so successful that a number of railroads requested the AAR to build cars for them. A total of 14 cars was built: 12 for the railroads and 2 for leasing service, all of which are still in regular testing service.

After these residual magnetic detector cars had been in service for some years, the Telewald Company developed another type of magnetic car, using what was termed "sustained-field" detection. On this car the pickups were carried behind the last magnet but close enough so that some of the leakage flux passed through the pickups. The test results and operating problems were very similar to those of the AAR's residual magnetic system. Two of these cars were built and are now owned by a western railroad.

The basic job of the AAR Detector Car Service is the development and construction of improved test equipment. This includes the installation of such equipment on railroad-owned cars and the Association's help in the operation and maintenance of this equipment. The construction and installation of new, much more powerful magnetizing equipment has recently been completed on all except one of the road cars.

Fig. 14 shows the results of all rail testing in the U. S. An all-time high record of both detected and service failures was recorded in 1943, after which there has been a very rapid decline of both. With the extensive testing program and the use of controlled cooled rail since 1938, the transverse fissure service failure has been practically eliminated. Most defects are now detected; these are various forms of progressive fractures such as shell fractures, engine burn fractures, detail fractures from gage check, etc.

For the future, another new system of detection is indicated: this is the ultrasonic method which is quite well developed in Germany. Fig. 15(A) shows the German Federal Railway ultrasonic detector car; (B), the ultrasonic recording camera and monitors; and (C), the ultrasonic crystal carriage and housings.

This car was built by the German Federal Railways and uses one vertical, two 30-degree and two 70-degree crystals to detect all types of rail defects. The car records photographically on a 35-millimeter paper tape. With this type of recording they are not limited to slow-speed testing but can operate up to 40 miles per hour very successfully. The one drawback with this type of recording is that the tape must be developed, dried, and then analyzed. After the tape is analyzed, word must be sent out to the road men advising them where the defects are located. This means a delay of about 5 days between the time the car runs over the track and when the defective rails are actually removed. The procedure is quite different from the testing procedure in this country. Here the railroads want to know immediately if the rail is defective because, if it is, they want to replace it before another train goes over it.

Fig. 16(A) illustrates an ultrasonic type of rail-flaw detector car developed by an engineer of the Atchison Topeka & Sante Fe (ATSF) railway system. Fig. 16(B) shows a detail of the Sante Fe car, which uses only one vertical probe on each rail, thus detecting horizontal-split heads, vertical-split heads and bolt-hole breaks in joints. This in-joint detection is its strong point because the ultrasonic method can test in the joint areas with no difficulty. The Sante Fe car is built in a "hi-rail" truck, i.e., it can operate on the track or on highways. By taking to the highway considerable time can be saved running to and from tests.

The Southern Pacific Railroad now has a similar car. The Sperry Company is also working on an ultrasonic car and is now operating, as an experimental car, one of their regular induction type cars on which is installed a vertical probe ultrasonic unit. This again is principally for the detection of cracks within the joint areas.

The AAR is now in the process of developing a miniaturized embodiment of the residual magnetism type of rail-flaw detector car. It will have flanged wheels for operation on the railway track, and also pneumatic tires by means of which the vehicle can be moved on highways. It is estimated that substantial economies can be effected by operation of a car of this type.

Fig. 17 shows the AAR *X201-X202* which has recently been rebuilt and on which is installed all of the latest type of test equipment. The improved magnetizing and test equipment has greatly improved both the speed of operation and the efficiency of the testing.

Discussion

Robert H. Webb (Sperry Rail Service, Danbury, Conn.): I should like to compliment Mr. Keevil on his description of the rail defect detector car. The purpose of this discussion is to elaborate on the developments made in the field of railway quality control by explaining some of the results of Sperry Rail Service research in this area. We believe that this research over the last 30 years has resulted in significant technical progress in the safety of American railroads.

As a point of departure from Mr. Keevil's historical discussion, the voltage drop method of rail defect detection should not be overlooked. While, indeed, the potential drop technique proved unsuccessful for dynamic testing due to the rail surface condition described, this technique has been employed by Sperry as well as other known induction and magnetic test cars over the past 30 years for the confirmation and classification of defects located by any of the dynamic systems. The size of transverse defects can be determined quite accurately by this measurement.

From the conception of a successful means for rail defect detection by Dr. Elmer A. Sperry in 1927 and 1928, our company has continuously invested in research and development toward the refinement of equipment and techniques for this vital service. H. C. Drake, Sperry research consultant, was recognized by the Franklin Institute award for his contributions to these developments. As the name implies, Sperry Rail Service provides on a contract basis a

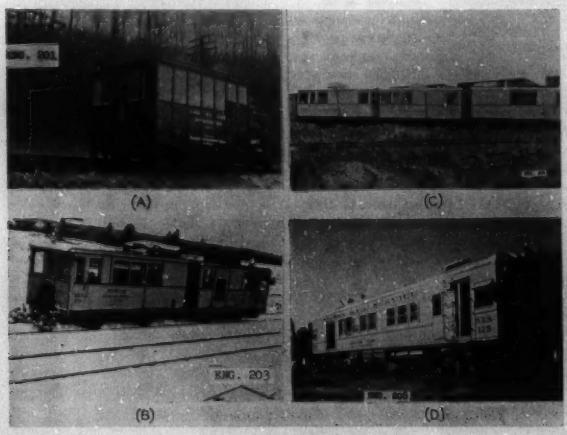


Fig. 18. Early designs of Sperry detector car. A—Single unit. B—Double unit. C—Triple unit. D—Current diesel electric-propelled car

Fig. 19. Current design of commercial test service

Fig. 20. Ultrasonic rail joint inspection car

Fig. 21. Special unit designed to negotiate severe subway grades

Fig. 22. Small, flexible tires housing various angle transducers, as sensing devices

quality control inspection; it covers currently over 150,000 miles of North American railways per year.

In briefly reviewing our progress toward rail safety I should like to show that the complexion of the Sperry detector car has changed significantly since 1928. Fig. 18 shows the development of early designs of equipment transported in small vehicles of single, double, and triple units; see Fig. 18(A), (B), and (C), respectively. The present diesel electric-propelled car, shown in (D), represents our increasing requirement for improved service has yielded more complex equipment to pace the expansion of the railroad industry and its attendant increased tonnage hauls through the years.

Providing the industry with a reliable commercial test service calls for rugged dependability which has been embodied in current design; see Fig. 19. Unitized construction, with transistor circuitry where practical, lends itself to optimum performance, ease of crew training and operation, and reduction of nonproductive time through quick change-out of the various equipment modules.

Industrial ultrasonic inspection equipment has been one of our standard product lines since 1944. In 1947 and 1948, experiments were conducted to determine the applicability of ultrasonic techniques to rail inspection. Particular areas of interest were the structurally joined rail ends. Cracks issuing from bolt holes as well as rail head and web separations eluded effective detection by the induction method because of the confused magnetic pattern set up in the joint area by the induction equipment. Research led to the development of the ultrasonic rail joint inspection car (Fig. 20) capable of detecting and evaluating bolt-hole cracks and head-web separations with uncanny accuracy. This service was made available in early 1950 and our present fleet stands at 13 cars, including three special units designed to negotiate the severe grades in subways; see Fig. 21.

We are at present installing ultrasonic rail detection equipment in our fleet of 15 induction cars for service; the first car was completed in July of this year. Resulting from 5 years' research and development, one element of this apparatus detects transverse defects occurring in the rail head, while another element detects bolt-hole cracks and head-web defects comparable to the service provided by the small ultrasonic car previously described. Small, flexible tires housing the various angle transducers are the

Fig. 23. New York transit system car, 1959

sensing devices for this system; see Fig. 22. These combination induction-ultrasonic cars will provide complete end-to-end rail testing with attendant rail safety not heretofore enjoyed.

As an adjunct to this newly developed equipment, a small portable ultrasonic hand test unit has been designed for defect confirmation and classification. This advance over the potential-drop method permits evaluation of defects under severe surface conditions caused by engine drive wheel burns, slivers, and what is known as shelly rail. The device has important implications, not only in the improved quality of test, but in increased production rates for the fleet.

It should be pointed out that, in 1954, Sperry made cross-licensing agreements with the Krautkramer Company in Germany, the company responsible for the advances in ultrasonic rail defect detection mentioned by the author. One of the principal purposes of this was a mutual exchange of technical information, a major concern being that of rail defect detection. In late 1957, Sperry representatives informed the American Railroad Engineers Association Rail Committee that we intended to bring such a car into the United States for demonstration to American railroads. At this writing, this car was expected to be delivered in August of 1959, to be placed into operation in the New York transit system; see Fig. 23.

Aluminum Bus for Shipboard Application

G. J. THOMPSON
MEMBER AIEE

S. H. BEHR
ASSOCIATE MEMBER AIEE

THERE is special interest in aluminum for shipboard use because of the savings realized in weight and because of shortages in copper that occur especially in times of national emergency when shipbuilding is at its peak. Although aluminum has been considered for cable conductors for shipboard use it is generally without over-all advantage because the larger conductor size results in a greater volume and weight of insulating and sheathing materials. These factors involving insulation are not imposed in applying aluminum to electrical bus. In utilizing aluminum bus for switchgear application problems arise that are mostly but not entirely confined to the making of satisfactory electric connections. The literature on the subject contains a large body of information. Much effort is devoted to coping with a particular environmental stress or material property considered to predominate in the reliability of a specific system. Some of the work stresses the advantage of a proprietary material or a particular method. Instances of controversy are attributed largely to differences of conditions of evaluation and as such appear best capable of resolution by re-evaluation on the same basis. The rather extreme conditions and their wide variety in shipboard service justify additional evaluation. A comprehensive experimental investigation was therefore carried out to determine the optimum means of utilizing aluminum bus in switchgear with the primary objective of long-term service reliability. Measurements taken on large numbers of specimens determine the extent to which different types of aluminum joined by various bolting methods, and with joint surfaces prepared with various joint compounds, are affected by prolonged load cycling, overload, weather-

Paper 59-654, recommended by the AIEE Marine Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19-21, 1959. Manuscript submitted January 19, 1959; made available for printing March 19, 1959.

G. J. THOMPSON and S. H. BEHR are with the New York Naval Shipyard, Brooklyn, N. Y.

The opinions or assertions contained in this paper are the private ones of the authors and are not to be construed as official or reflecting the views of the Naval Service at large.

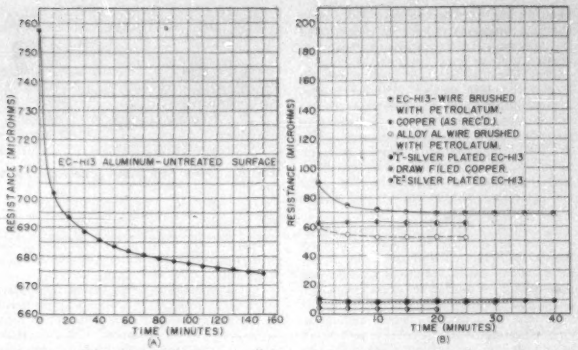


Fig. 1. A—Joint resistance at 3,000 psi. B—For copper and prepared aluminum bars

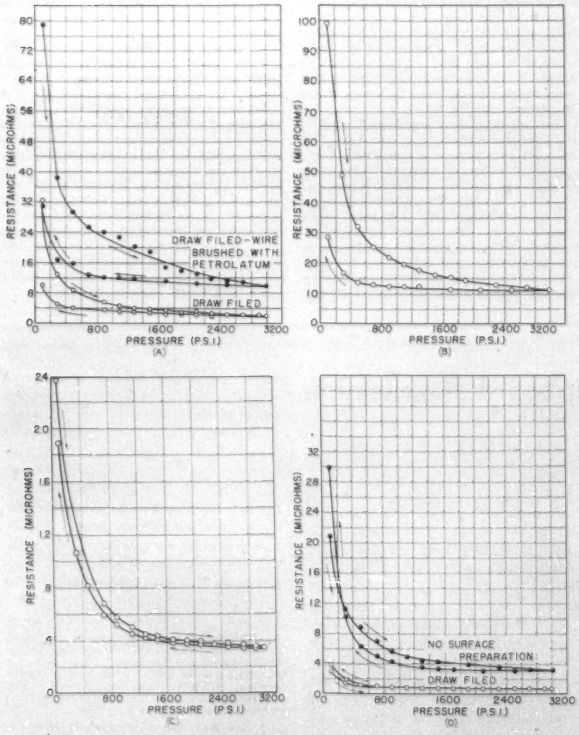


Fig. 2. A—Joint resistance versus pressure: EC-H13 aluminum. B—EC-H13 aluminum wire-brushed with joint compound. C—EC-H13 aluminum silver-plated surface. D—Copper bars

ing, vibration, and high impact shock; similarly tested is the extent to which silver platings are affected. Measurements on an actual marine installation after 20 years of service are correlated with the results of experimental laboratory work. In addition, information is provided which is useful in the design of bus supports needed to withstand fault currents.

Measurements on Stacked Bars

Measurements on specimens of typical joints were preceded by studies of the variation of joint resistance under carefully controlled mating pressures. This was done on stacks of bar material by the method previously employed by Denault[1] and others.[2] Electrical aluminum bus of two grades of hardness was used. The softer grade is commercially known as EC-H13 aluminum bus conductor. The harder grade is an aluminum magnesium silicide alloy bus conductor[3] which is hereinafter termed "alloy bus." The stacked bar technique consists essentially of measuring the series resistance of a number of flat specimens held together at various steps of constant pressure. Each setup consisted of eleven $1^1/_2$- by 1- by 1/4-inch specimens of electrical aluminum of the same grade with the blocks arranged to cross each other at right angles so as to form ten mating surfaces of one square inch area each. Pressure on the stacked specimens was maintained by a well-sealed hydraulic press and the resistance was measured by a Kelvin double bridge using a current of 10 amperes to minimize heating effects. The value measured included the resistance of the aluminum material as well as the joint resistance. The resistance of the body of aluminum in the stack is between 2.7 and 3 microhms. Figs. 1 and 2 accordingly show the average resistance of one joint and the material in one block.

An indication of the extent of heating effect of current on resistance measurement was gained from measurements of joint resistance on stacks of uncleaned aluminum bars while the current was increased from 10 to 50 amperes. Joint pressures used varied from 2,000 to 8,000 psi (pounds per square inch) in 1,000-psi increments. Below 4,000 psi there was a progressive increase in resistance with current. The rate at which resistance increased with current was greater at lower than at higher pressures. At 4,000 psi the resistance rose less than 1.0% as the current was increased from 10 to 50 amperes; at 2,000 psi the rise was about 4%. Although

measurements were made rapidly to minimize temperature rise, the increased resistance is nevertheless attributed to an undetected temperature rise of the actual contact points which are of very small area[4] and which exist in smaller numbers at lower joint pressures than at higher. As a result of this exploratory work subsequent measurements were made at 10 amperes, which was also a convenient current for the Kelvin bridge employed.

Measurements made on stacked blocks of untreated EC-H13 aluminum under constant pressure show a change of resistance with time. Joint resistance dropped 7% of initial value 10 minutes after applying 3,000-psi joint pressure and continued to diminish, tapering off to an 11% drop in 2½ hours, as shown in Fig. 1(A). Additional evidence shows substantial stabilization in 3 or 4 hours. This is attributed to creep of the aluminum. The deformation results in the production of an increase in the number and size of actual contact areas within the mating surface, as the oxide layer is broken through, thus progressively lowering the joint resistance until further yield of the metal has ceased.[4] This demonstrates the necessity of stipulating the time interval between application of the pressure and the making of resistance measurements. Subsequent measurements indicated that resistance stabilization could be achieved earlier with less degree of change and at lower values when higher pressures were employed. Measurements of this nature, as shown in Fig. 1(B) on draw-filed copper bars and on silver-plated[5] aluminum bars indicate very little change of resistance with time. Aluminum bars, that had been brushed with and coated with petrolatum, exhibited a lowering of resistance with time, becoming substantially stabilized after 10 minutes. Where applicable thereafter, the resistances reported are those measured 10 minutes after application of pressure. During the making of measurements at 1,000 psi it was observed, particularly for low resistance, that joint resistance varied as much as four to one between different adjacent blocks within the stack. The variability existed to a greater extent at lower pressures. In another experiment the blocks were rearranged within the stack and the resistance of the stack varied about 20% in six trials. Such variability is due principally to joint-mating surface condition and to a much lesser extent by instrument limitation. At stack resistances of over 5 microhms the variability is almost entirely due to joint surface variation. The figures reported herein are for an average of ten joints, which tends to minimize the effect of this variability.

The effect of relaxation of pressure on joint resistance is shown in Fig. 2(A), which is the result of measurements on stacked bars of draw-filed EC-H13 aluminum. Joints made on petrolatum-coated aluminum bars are shown in this figure to exhibit only about a 25% rise in resistance when joint pressure is relaxed by as much as 75%. It is further observed that the slope of the resistance versus pressure curve is significantly greater when the pressure is being increased than when it is being decreased. This was found to be generally characteristic of joints made with unplated EC-H13, the relatively softer grade of aluminum. This is demonstrated by Fig. 2(B). Virtually no difference was found between slopes for ascending and descending pressures for compound brushed joints of the harder alloy aluminum, copper, or silver-plated aluminum of either of the two grades of hardness. Fig. 2(C) illustrates this effect. Fig. 2(A) shows the insulating effect of petrolatum on joints made on EC-H13 aluminum. The film insulating effect was found so much more serious with silicone grease as to preclude its consideration as an inhibitor. Further measurements showed no significant differences in joint resistances between silver-plated EC-H13 surfaces whether they were coated or not coated with petrolatum. Repeating this type of test on unplated EC-H13 showed an increase in joint resistance of about five times when the surfaces of the same blocks were coated with petrolatum. Electrodeposited silver has a Vickers hardness pyramid number of 100 as compared to 38 for a specimen of EC-H13. This suggests that the "high points" on the harder silver mating surface may be more effective in shearing through the petrolatum film than the high points of the softer EC-H13 aluminum. Figs. 2(C) and (D) both show relatively low resistance and, in addition, show a narrow spread between curves for increasing and decreasing pressure. Figs. 2(B) and (C) indicate that the resistance of silver-plated joints in the region of 100 to 500 psi is between one tenth and one fiftieth that of compounded joints, and that at these low pressures the silvered joint resistance is less than is generally exhibited by compounded joints at 1,800 psi. This means that a joint of acceptably low resistance can be made with silver-plated bars at a fraction of the bolting torque required for compounded joints. While this does not furnish a basis for condoning the use of a loose or mechanically inferior joint, it does suggest the extra margin of electrical integrity offered by the silver-plated joint. Mechanical disturbance of a joint, incidentally, may be considered to produce no deleterious effect unless it progressively loosens the joint as motion between mating surfaces introduces a wiping action which tends to reduce resistance as described by Xenis.[6] The measurements of resistance change with decrease of joint pressure tend to show that retightening connections on aluminum bus does not decrease joint resistance to the extent that measurement of mechanical relaxation of joint pressure may seem to indicate.

A comparison of the resistance of various types of joints may be made from the results of measurements on stacked bars. Table I is an excerpt of resistance of the various joints at 1,800 psi, a joint pressure approximating that exerted by a single

Table I. Joint Resistance at 1,800 PSI, Microhms PSI

Bus Material	Contact Surface Preparation	Manufacturer Designation	Resistance
EC-H13	silver-plated	E	0.4
		F	0.4
		G	0.4
		H	0.7
		I	1.3
		J	0.6
		*	0.6
Alloy bus	silver-plated	K	0.6
Copper	no preparation		3.0
Copper	draw-filed		0.6
EC-H13	wire-brushed with compound	L	11.0
		M	1.0
		N	6.1
		O	3.0
		P	12.5
	wire-brushed with no compound		3.0
	wire-brushed with petrolatum		25.0
	wire-brushed with silicone grease		400
Alloy bus	wire-brushed with petrolatum		10.0
EC-H13	no preparation		310
Alloy bus	no preparation		110

* Matlab Navshipydnyk.

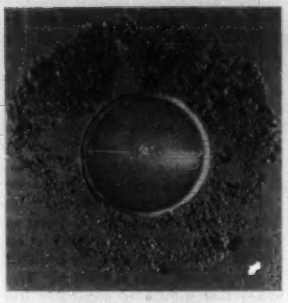


Fig. 3. Pitted contact area

1/2-inch steel bolt torqued to 50 pound-feet on a 2-inch lapped, 2-inch wide bus. Table I shows that there is very little variability of resistance of joints silver-plated by different plating processors. The resistance of silver-plated joints was several times lower than was general for compound treated joints. Silver-plated aluminum produced about the same joint resistance as draw-filed copper. The figures shown in this table may also be compared to the resistances of bus joints in an old installation made by the technique of abraiding the joint surfaces under an inhibiting compound. The installation had been in service aboard ship for about 20 years when the measurements were made. Resistances of 20-year old joints chosen at random were found to be between 0.5 and 300 microhms per square inch. By comparison with the figures in Table I these old connections obviously ranged in resistance from that of the lowest practicably attainable to that resulting from total lack of contact surface preparation.

Cycling Tests

When the effects on joint resistance of the creep of the material and the type of surface preparation had been observed, an operational evaluation in the form of load cycling was undertaken. The load cycle[2] consisted of sufficient current to obtain a 35 C (degrees centigrade) bus temperature rise above a 50 C ambient for 2 hours followed by a shutdown period of cooling at about 25 C for 1 hour. All specimens were subjected to 1,000 such cycles followed by 10 cycles at 125%, 150%, 175%, and 200% of the previous load current. Each specimen consisted of a 3-foot section of 2- by 1/4-inch aluminum bus. Each group of specimens consisted of four 3-foot sections connected in series with three identical single-bolt joints. All groups were connected in series to a low-voltage power source. These were mounted in the same plane in a controlled ambient chamber, each group being isolated from the radiant heat of the adjacent group by baffle plates of thermal insulation. Materials constituting the specimens were as follows:

EC-H13 aluminum bus: 61% International Annealed Copper Standard, tensile strength 17,000 psi.

Alloy bus, aluminum magnesium silicide: 55% (copper standard), tensile strength 29,000 psi, represented by three different manufacturers.

Silver-plated aluminum bus: zincate plating process as produced by six different manufacturers.

Copper bus, electrical grade.

Joint compounds, five different manufacturers.

Aluminum alloy bolts and nuts coated with waxy lubricant: 5/8 inch, 11 threads per inch with 1¹/₄-inch-outside-diameter aluminum washers.

Galvanized steel bolts and nuts: 1/2-inch, 13 threads per inch with galvanized steel washers, Belleville type spring washers rated at 6,600 pounds flattening force.

Steel spring lock washers; for 1/2-inch bolt, medium size.

Types of joint mating surfaces included the following:

As received (no preparation).
Draw-filed.
Wire-brushed.
Wire-brush through joint compound.
Silver-plated.

Bolting methods at 50-pound-foot tightening torque included the following:

Galvanized steel bolt and nut with flat steel washers.

Galvanized steel bolt and nut with flat steel and split spring lock washer.

Galvanized steel bolt and nut with flat steel and Belleville-type spring washer.

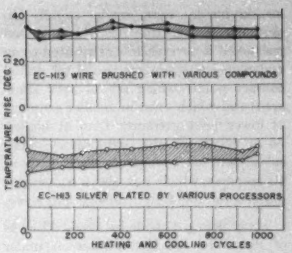


Fig. 4. Full-load temperature rise of joints

Aluminum bolt and nut with two flat aluminum washers.

All joints were made with a 2-inch overlap and held with a single bolt. Two thermocouples were cemented on each bus connection for temperature measurements. Millivolt drops were measured between points 4 inches apart across each connection. Load cycling, together with ambient temperature cycling of the chamber was automatically controlled, and temperatures were automatically recorded. D-c loading was substituted for a-c loading at intervals to facilitate measurement of millivolt drops of the connections.

To study the behavior of poorly made connections, the effect of load cycling was observed on joints made without surface treatment. In a few instances these joints exhibited small temperature rise and low resistance, showing that it may be occasionally possible to secure an apparently satisfactory joint without proper preparation. This circumstance is attributed to a favorable, although haphazard, degree of abrasion of the oxide film during the assembly of the connection. Most of each of the other joints made without surface preparation exhibited wide reversals in trends of resist-

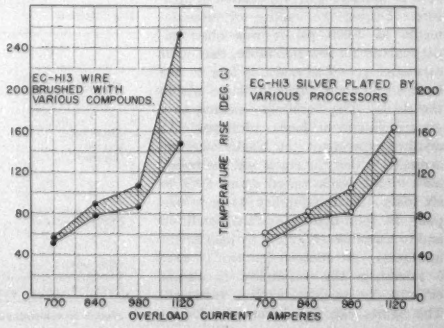


Fig. 5. Overload temperature rise of joints

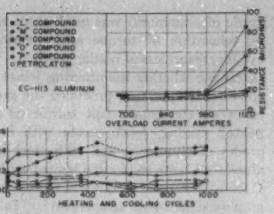
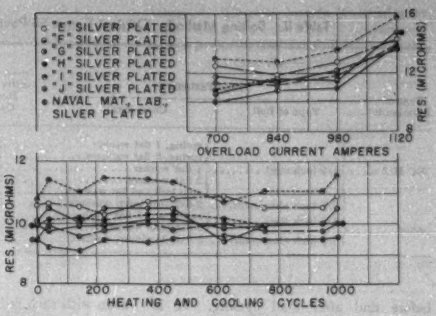


Fig. 6. Resistance of wire-brushed compounded joints

Fig. 7. Resistance of silver-plated joints



ance and temperature changes during cycling, which may indicate mechanical disturbance caused by load cycling. After 1,000 cycles of loading as described, temperature rises ranged from 40 to 175 C and joint resistances ranged from 20 to 400 microhms. Load cycling of these poorly made joints synthesized the so-called "runaway" condition caused by the interaction of resistance and temperature which progressively deteriorates the joint, culminating in destructive temperature rise. One mating surface of such a joint is shown in Fig. 3. Joints like this were consistently found to produce higher temperature on a-c than on d-c loading. Arcing, particularly at the start of the load cycle, led to the disclosure that the welds which were being formed at the interface were less substantial if they had been formed by alternating current than by direct current. This resulted in a higher resistance during the passage of alternating current.

The extent to which temperature rise and resistance is reduced by the various methods of preventing the formation of oxide film is shown by the results of load-cycling well-prepared joints. Temperature rise ranged from 25 to 38 C and resistance ranged from 9 to 15 microhms. Fig. 4 shows that temperature rises of silver-plated and compound-prepared joints fastened with a steel bolt, flat washer, and dished spring washer are of the same order. Overload cycling shown in Fig. 5 shows that the silver-plated joint runs only slightly cooler at the most extreme overload employed. Figs. 6 and 7 were drawn with an expanded ordinate scale for the purpose of maintaining the identity of the results for each type of joint. This has the effect of exaggerating the swings in resistance. It is observed that the compounded joints under cycling are generally about 15% higher in resistance than the silver-plated joints and that this difference tends to widen as overload cycling is applied. Furthermore, both the load cycling and the overload cycling show that compound-prepared and silver-plated joints are stable. Cycling tests were made on a wide variety of combinations of bolting methods applied to both EC-H13 and alloy aluminum bus. The results of only a few of these combinations are shown in Figs. 8 and 9. All joints appear stable and differences between joint temperatures and resistance are not wide enough to rule out any material or method. The simplest form of connection, that of the steel bolt and nut with flat washers, when used on alloy aluminum bus, maintained a relatively low resistance and temperature rise throughout load and overload cycling. The cycling results indicate no significant advantage attributable to any of the various forms of springs used on alloy aluminum bus.

Joint Pressure Relaxation

Neither joint temperature nor resistance change during the load cycling employed is a completely reliable measure of the extent of the joint pressure relaxation which can be expected with aluminum bus. Earlier it was demonstrated with stacked aluminum bars that joint resistance change is not a direct measure of joint pressure change. Therefore, to find joint pressure relaxation, torque measurements were made on connection bolts

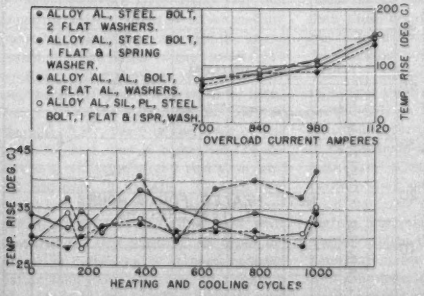

ALLOY AL., STEEL BOLT, 2 FLAT WASHERS.
ALLOY AL., STEEL BOLT, 1 FLAT & 1 SPRING WASHER.
ALLOY AL., AL., BOLT, 2 FLAT AL. WASHERS.
ALLOY AL., SIL. PL., STEEL BOLT, 1 FLAT & 1 SPR. WASH.
TEMP. RISE (DEG. C)
OVERLOAD CURRENT AMPERES
HEATING AND COOLING CYCLES

Fig. 8. Temperature rise of joints: various bolting methods

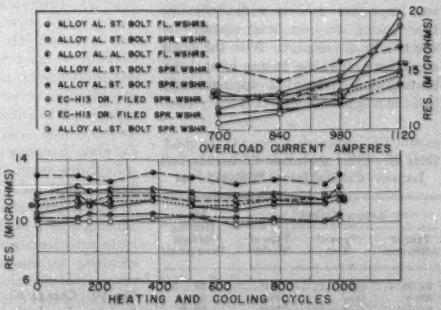

ALLOY AL. ST. BOLT FL. WSHRS.
ALLOY AL. ST. BOLT SPR. WSHR.
ALLOY AL. AL. BOLT FL. WSHRS.
ALLOY AL. ST. BOLT SPR. WSHR.
ALLOY AL. ST. BOLT SPR. WSHR.
EC-H13 DR. FILED SPR. WSHR.
EC-H13 DR. FILED SPR. WSHR.
ALLOY AL. ST. BOLT SPR. WSHR.
RES. (MICROHMS)
OVERLOAD CURRENT AMPERES
HEATING AND COOLING CYCLES

Fig. 9. Joint resistance: various bolting methods

Table II. Bolting Method: Original Torque 50 Pound-Feet

Type of Aluminum	Joint Fasteners: Type of Bolt	Joint Fasteners: Components	Torque, Pound-Feet: After 1,000 Heating Cycles	Torque, Pound-Feet: After 1,000 Heating Cycles and Overload
EC-H13	1/2-inch steel	spring, 1 flat washer	29	16
		spring, 2 flat washers	20	4
		1 flat washer	45	10
		2 flat washers	27	2
		split lock and flat washers	20	3
	5/8-inch aluminum	2 flat aluminum washers	50	32
Alloy	1/2-inch steel	spring and 1 flat washer		77
		2 flat steel washers		51
	5/8-inch aluminum	2 flat aluminum washers		102

before and after load cycling. It is recognized that torque measurement includes friction as well as pressure components which are related and not readily separable. It is assumed, however, that friction will not decrease unless pressure decreases, so that diminished torque can be attributed to diminished pressure under the bolt head. As shown in Table II, there are instances of increases in torque above the originally applied torque. These inconsistencies, which are due to the nature and the material of thread surfaces, to the lubricant, and to the nature of the nut face, have been investigated and described by Pickel.[8] Despite such limitations, torque measurements were found to coincide with other indications of pressure relaxation. Table II shows that the cycling used to effect changes in joint resistance and temperature rise caused reductions in torque whenever EC-H13, the softer grade aluminum, was used with steel bolts; this was true regardless of attempts to compensate for unequal coefficients of thermal expansion. Neither the Belleville type of spring washers nor split-lock washers showed any superiority over flat washers in preventing loss of joint pressure as measured by torque after heat cycling. Spreading the bolting pressure over a wider area by using two large, flat washers with Belleville-type spring washers was equally ineffective in this respect. With the single exception noted, the torque after 1,000 heating cycles relaxed to approximately half the original values. It is likely that, even with such reduction in contact pressure, satisfactory joint resistance could be maintained, because joint resistance increases at a very low rate while the joint is undergoing considerable reduction in pressure, as has been previously shown in studies on stacked bars. After overload cycling, the torque on all steel bolts joining EC-H13 grade aluminum relaxed to 10% or 20% of the original. The corresponding contact pressure produces a joint which is likely to have excessive resistance and to be considered unreliable. The use of the aluminum alloy bolt (the next size larger than steel, according to current commercial practice), with an aluminum nut and flat aluminum washers, produced no torque relaxation after 1,000 heat cycles and a reduction of about 30% of the original after overload cycling. Where the alloy bus (which is the harder grade of aluminum) was used, there was no relaxation of torque after overload cycling, regardless of bolting method. To determine whether maintaining joints at elevated temperature produced the same effect on connection bolt torque as repeated cycling, joints made with both grades of aluminum and with steel and aluminum bolts were oven-heated to 150 C for 16 hours and torque was remeasured afterward at room temperature. Table III gives the results of the average of five specimens for each connection. These results show that connection bolt torque responds to a single temperature change between room temperature and 150 C in a similar manner to repeated heat cycling. Where the harder grade aluminum (alloy bus) is used no relaxation of torque is indicated, even without spring washers, as contrasted with the results where EC-H13, the softer grade aluminum, is used. Torque measurements, taken as an indication of contact pressure, show that the choice of the type of aluminum bus as well as the bolting method depends on the temperature and the temperature range of operation. Joints made with the harder grade aluminum appear capable of withstanding a wider temperature variation than those of the softer grade. Fig. 10 shows the effect of temperature on relative hardness of EC-H13 and alloy bus aluminum. This figure is based on short time exposure to elevated temperatures. Longer periods of time at lower

Table III. Effect of Heat on Connection Bolt Torque: Original Torque 50 Pound-Feet

Bolting Method: Type of Aluminum	Bolting Method: Type of Bolt	Bolting Method: Type of Washer	After Heating Cycles, Pound-Feet
EC-H13	Steel	Flat	25
Alloy	Steel	Flat	55
EC-H13	Aluminum	Flat	62.5
Alloy	Aluminum	Flat	93.0

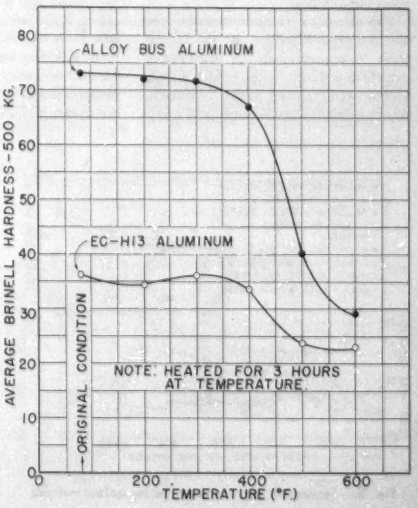


Fig. 10. Change of hardness in aluminum

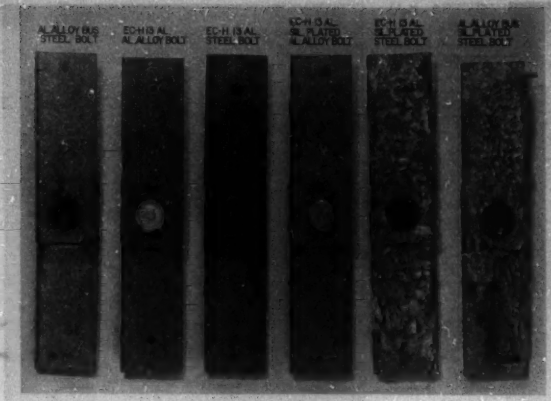

AL ALLOY BUS STEEL BOLT
EC-H13 AL AL ALLOY BOLT
EC-H 13 AL STEEL BOLT
EC-H 13 AL SIL PLATED AL ALLOY BOLT
EC-H 13 AL SIL PLATED STEEL BOLT
AL ALLOY BUS SIL PLATED STEEL BOLT

Fig. 11. Corrosion test specimens

temperatures may be expected to produce a similar effect.

Corrosion Effects

Short specimens of aluminum bus with lapped joint connections representing the grades of aluminum, bolting methods, and types of contact surface preparation employed in the load cycling tests, were assembled and mounted on a weathering rack facing due South and also facing the ocean. The rack is at a seaside location beyond the immediate New York City urban area which provides wide temperature variation with time of day and season, salt-laden air and wind-driven sand and rain. This is an extreme condition of exposure. Fig. 11 shows the severity of corrosion displayed by typical joint specimens exposed to this weathering for about 1½ years. Periodic measurements of joint resistance were made over the exposure period. Table IV shows median values of the results of several of the joints tested. Examination of the detailed results shows that the joint resistance does not change uniformly with exposure time. Joints exhibiting resistance increases at the end of 1½ years in the order of ten times original show a higher rate of increase in resistance after 1 year than during the first 12 months. Cleaned dry joints held with aluminum bolts show less resistance change than those held with steel bolts. Dry joints of all types increase in resistance to a much greater extent than joints protected by a coating of joint compound. The greater joint deterioration observed when steel bolts and washers were used is partly attributed to the physical weakening of the joint by the partial disintegration of the steel. The wax-coated aluminum bolts and washers suffered no appreciable deterioration even when fastened to silver-plated bus. These results generally confirmed those of Connor and Wilson,[9] indicating the superior stability of silver-plated aluminum joints.

Short-Circuit Tests

The difference in mechanical properties between present electrical grade aluminum alloys and copper suggests the need of bus supports and spacers especially suitable for aluminum bus work. Space conservation aboard ship generally requires closer spacing between phases than is customary in industrial practice. This in turn produces greater electromagnetic forces for the same short-circuit currents than those occurring with wider spaced bus. To determine the extent of bus deformation, bus structures of EC-H13 and alloy bus were set up and subjected to 3-phase short circuits corresponding to the short-circuit capability of the installation contemplated. Each phase consisted of two 3- by 1/4-inch aluminum bars 1/4 inch apart. Center-to-center spacing between phases was 3 inches and 7/8-inch glass melamine supports were spaced at various intervals of from 6 to 24 inches. Table V shows the maximum permanent deformation for each spacing produced by a 3-phase 60-cycle 100,000-ampere short-circuit current of 5 cycles duration. Figs. 12 and 13 show the bus structures

Table IV. Effect of Seaside Weathering on Joint Resistance of EC-H13 Aluminum 1/2-inch Galvanized Steel Bolt, Belleville, and One Flat Steel Washer

Type of Surface Preparation	Resistance, Microhms					
	Results, No. of Months Later					
	Initial	3	6	9	12	15
No preparation; no protective compound	15	45	65	72	70	88
Silver-plated joints; no protective compound	5	10	10	10	19	60
Compounded joints	7	9	7	7	7	9

before and after imposing the fault current. It is recognized that the measurements of deformation shown in Table V are taken from bus distortions resulting not only from the magnitude of current and bus spacing but depending on the time of current interruption. For this reason the action was observed by means of high-speed motion pictures taken at 2,000 frames per second. These pictures showed actual contact and arcing between sections of bus bars of adjacent phases which later appeared completely separated, following the current interruption. Table V is to be taken only as an approximation of what can be expected as a result of such a short circuit.

Vibration and Shock Tests

Vibration, which is generally more prevalent and severe aboard ship than in industrial plants, is a condition that was considered as having possible effect on the reliability of bolted aluminum bus bar connections. Accordingly, specimens of 2- by 1/4-inch aluminum bar connection were mounted on a vibration machine table with 15 inches between supports, and with the bolted connection mid-way between supports. The specimen was vibrated in the planes of both major and minor axes with an amplitude of 1/16 inch (total excursion of 1/8 inch) for 2 hours. The frequency of vibration was varied

Table V. Maximum Permanent Deformation in Inches of Bus Following 3-Phase 60-Cycle 100,000-Ampere Short Circuit

Type of Aluminum	Support Spacing, Inches, Center to Center						
	6	8	12	15	18	21	24
EC-H13	0	3/4	3/4	3½	1¼	support broken	
Alloy	0	0	1/4	¾	2½	5½	

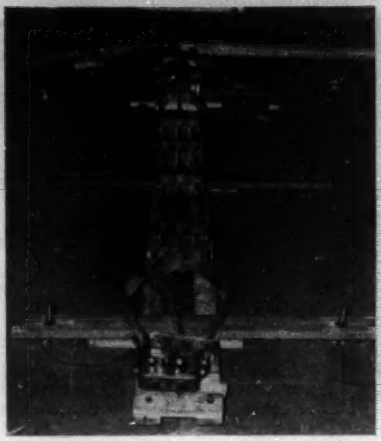


Fig. 12 (left). EC-H13 bus after 3-phase 100,000-ampere short circuit

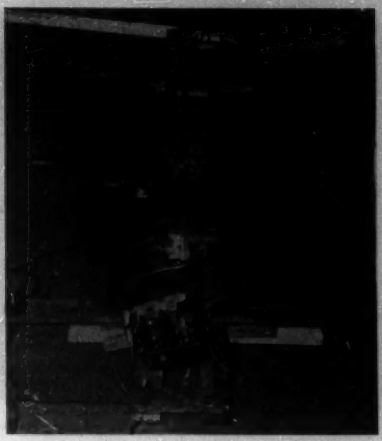


Fig. 13 (right). Alloy bus after 3-phase 100,000-ampere short circuit

from 10 to 55 cycles per second and back to 10 cycles per second each minute. The electrical resistance of the joint and the bolt torque of each specimen were measured before and after the 2-hour vibration period. Four specimens of each connection were subjected to vibration. Aluminum bolts and steel bolts with and without spring washers were used in connecting both grades of aluminum. There was no evidence of loosening after vibration. Joint resistance decreased in all cases, indicating that vibration may have produced sufficient additional flow to increase the number of joint interface contact points, but with too small a loss of contact pressure to be detected by torque measurements.

There was some concern about aluminum bolts breaking under high-impact shock because of the general practice of tightening to within 80% of the ultimate strength. Two specimens of 2- by 1/4-inch EC-H13 aluminum bus joined with 5/8-inch aluminum bolts were bolted at the ends to the high-impact shock machine specified in reference 10. The specimens were mounted with 15 inches between supports, the joints being positioned mid-way between supports. The 2-inch surface of the bus faced the direction of impact blow. The connections were subjected to five 2,000-foot-pound blows. There was no significant change in connection bolt torque and no permanent deformation of the bus bars.

Conclusions

The choice of aluminum bar material, contact surface preparation, bolting method, and joint protection depends upon the nature and severity of the service for which the installation is intended. It is concluded that an aluminum bus installation in a protected inboard location aboard ship should include the following features:

1. The use of aluminum magnesium silicide alloy, which is preferable to the softer EC-H13 aluminum as bus conductor because bolted joints made with it are more electrically stable and it has greater resistance to deformation under fault currents within a bus structure.

2. Silver plating of the entire bus by the zincate process after forming and punching, which is preferable to compound brushed joint preparation because the joint on silver-plated bus has lower resistance. The resistance is lower than that of a compounded joint, even when the bolting pressure is but a fraction of that required for a compounded joint, thereby ensuring a greater margin of electrical integrity. In addition, the quality of connection is obviously more uniformly controlled, with resulting improvement in reliability.

3. Coating of the silver-plated joint surfaces with a joint compound before assembly, which is an effective measure of insurance against long-time corrosion effects.

4. The use of one or more steel bolts and nuts with flat washers, which is sufficient to maintain effective joint pressures when it is used with aluminum alloy bus. Retightening of bolts is generally unnecessary.

5. Bus support and bracing against fault currents required for alloy aluminum bus, which is substantially the same as required for copper bus of equal current-carrying capacity.

The investigation also furnished information applicable to the operation of the bus at over 85 C. An added measure of connection reliability may be secured for such operation by the use of aluminum bolts and nuts. The substantially equal coefficients of thermal expansion of aluminum bolt and bus materials will prevent metal flow to which aluminum bus material is increasingly susceptible at elevated temperatures. Torque-limiting wrenches should be used on aluminum bolts to prevent overstressing. The lack of absolute imperviousness of silver platings and the acceleration of oxidation rate with increase in temperature place further reliance on a joint compound coating with good retention ability. The investigation shows alloy bus to be capable of operation at somewhat higher temperature than EC-H13 but its limit as concerns softening with temperature and time has not been determined in this investigation.

While not primarily intended to furnish information for performance under weathering, the investigation showed that more than a year of severe atmospheric weathering was required to establish significant results. Because of the difficulty of maintaining protection for steel connecting bolts, it is concluded that aluminum bolts applied with a torque-limiting wrench will produce a more reliable joint. As all dry joints deteriorated when exposed to weathering and as joint compound coatings inhibited this action, it is concluded that the use of a compound is necessary for such service.

References

1. Electric Contacts of Bus Bar Joints, C. L. Denault, *Electric Journal*, East Pittsburgh, Pa., vol. 30, 1933, pp. 281–83.

2. Alcoa Aluminum Conductor Handbook. Aluminum Company of America, Pittsburgh, Pa., 1957, pp. 102–06.

3. Use and Properties of Extruded High-Strength Aluminum for Electric Bus Conductors. W. Swinney, C. L. Carlson. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 75, June 1956, pp. 449–53.

4. Electric Contacts (book). Ragnar Holm. Gebers, Stockholm, Sweden, 1946, pp. 7–9, p. 86, p. 159.

5. Electroplating of Aluminum Alloys. *Bulletin no. 7*, Aluminum Company of America, June 1954.

6. Fundamental Problems Encountered in Aluminum Connections. C. P. Xenis. "Symposium on the Use of Aluminum for Insulated Conductors," *AIEE Special Publication S-56*, Oct. 1953, pp. 20–23.

7. Effect of Heat Cycling on Bolted Aluminum Bus Joints. H. E. House, N. T. Bond. "Proceedings of the Conference on the Electrical Utilization of Aluminum," *AIEE Special Publication T-72*, Mar. 1955, pp. 94–98.

8. Tightening Characteristics of Nut, W. F. Pickel. *Product Engineering*, New York, N. Y., Jan. 1949, pp. 98–102.

9. Performance of Electrical Joints Utilizing New Silver Coating on Aluminum Conductors, T. J. Connor, W. R. Wilson. *AIEE Transactions*, pt. III, vol. 72, Aug. 1953, pp. 702–12.

10. *Military Specification MIL-S-901B*, U. S. Navy, Washington, D. C., 1954.

11. Mechanical Properties of Aluminum Electric Bus. G. W. Stickley, C. O. Smith. *AIEE Transactions*, pt. III-A, vol. 73, Apr. 1954 pp. 100–06.

12. The Effect of Surface Preparation on the Electrical Resistance and Temperature Rise of Aluminum Bus Conductor Joints. *Report no. 5603-12*, Westinghouse Materials Engineering Department, Pittsburgh, Pa.

13. Aluminum in Heavy Current Conductors, William Deans. *AIEE Transactions*, pt. III, vol. 74, Dec. 1955, pp. 1192–1200.

14. Aluminum for Marine Switchgear, H. F. Harvey, Jr., E. J. Dawson. *Ibid.*, vol. 75, Apr. 1956, pp. 134–42.

15. Contact Resistance—The Contribution of Nonuniform Current Flow, W. B. Kouwenhoven, W. T. Sackett, Jr. *Ibid.*, pt. I, vol. 70, 1951, pp. 791–95.

Discussion

N. T. Bond (Alcoa Research Laboratories, Massena, N. Y.): It is encouraging to those interested in using aluminum bus conductors that investigations as comprehensive as this are being reported. However, it is unfortunate that no more distinction is made between the several proprietary joint compounds than a difference in the letter by which they are identified in this investigation.

There are vast differences between joint compounds both in the job they are intended to do and in the manner in which they do it. Some joint compounds are "grease-like" materials which serve primarily as protective sealers. It is for this particular function that a joint compound is recommended, in conclusion 3, for joints in silver-plated conductor. Other joint compounds contain particles which deform the contact surfaces. Still other joint compounds contain active chemicals which are effective in bringing about low contact resistance. Alcoa no. 2 electrical joint compound is an example of the last-mentioned type. Each of the joint compounds has as one of its functions that of protecting the joint against environmental corrosion.

It is pointed out that "Figs. 2(B) and (C) indicate that the resistance of silver-plated joints in the region of 100 to 500 psi is between one tenth and one fiftieth that of compounded joints . . ." but the type of joint compound used is not indicated. In addition, conclusion 2 states that "The resistance [of a silver-plated joint] is lower than that of a compounded joint, even when the bolting pressure is but a fraction of that required for a compounded joint . . ." Fig. 2(B) gives the resistance of a joint which was wire-brushed with joint compound as 14.5 microhms at 1,800 psi, whereas Table I gives the resistance of a joint wire-brushed with compound *M* as 1.0 microhms at 1,800 psi. In fact, the latter resistance is lower than was obtained from the joint silver-plated by manufacturer *I*, which was 1.3 microhms; see Table I. It would appear that the statement regarding Figs. 2(B) and (C) and that of conclusion 2 should be qualified or restricted to certain joint compounds. It is important that a distinction be made between different joint compounds, or at least between their different types.

In the discussion of corrosion effects it is pointed out that "Cleaned dry joints held with aluminum bolts show less resistance change than those held with steel bolts." It is also stated that "The greater joint deterioration observed when steel bolts and washers were used is partly attributed to the physical weakening of the joint by the partial disintegration of the steel." It is worth mentioning that the differential expansion of the steel and aluminum when subjected to normal temperature variations would cause the clamping force to vary and would allow corrosion to destroy some of the contact spots which would not have been destroyed if the clamping force had been constant.

Because the purpose of the investigations was ". . . to determine the optimum means of utilizing aluminum bus in switchgear with the primary objective of long term service reliability," it would appear that aluminum bolts should be the primary recommendation with the use of steel fasteners subordinated.

G. J. Thompson and **S. H. Behr**: We thank Mr. Bond for his comments. Although we agree that identification of joint compounds rather than the use of code letters would be more informative, the work which furnished the basis for the paper was conducted under authorization of the Navy, and identification of competitive manufacturers was deliberately avoided as a matter of policy. However, in Figs. 4 and 6 it is demonstrated that the temperature rise and the joint resistance for full load current during and after 1,000 heating and cooling cycles are of the same order regardless of the type of proprietary joint compound used.

Concerning Mr. Bond's comment on Figs. 2(B) and (C), the curves shown were obtained by the stacked block method. Although resistance measurements for all joints compounds were made on aluminum blocks, Fig. 2(B) was selected as representative of curve shape and was code *L* compound: Table I shows 11 microhms as release resistance. Compound *M* showed a contact resistance comparable to silver-plated connections but it was slightly higher than all zincate process plated specimens. Manufacturer *I* would not divulge his method of plating which was other than the zincate process. Further to qualify conclusion 2, the authors believe that satisfactory aluminum bus connections can be prepared using joint compounds such as codes *M*, *N*, or *O* providing careful workmanship and sufficient working space is available.

Shipboard bus installations are normally crowded, often making it impractical to prepare reliable wire-brushed joints, especially on multibar connections. An analysis of several shipboard installations substantiated by voltage drop measurements made on a 20-year-old vessel with wire-brushed aluminum connections proved this latter statement. The resistance of many of these bus connections were of the same order as that of well-prepared new joints, yet others in inaccessible locations were of a value that would cause severe overheating had full-load current been applied. For this reason the authors recommend silver-plated connections for shipboard (or similar) installations so that connecting methods similar to those used for copper will be applicable.

With reference to Mr. Bond's question on what joint compound was used on the silver-plated blocks tested, represented by Fig. 2(C), these blocks were uncoated to facilitate comparison of silver platings by various processors. The application of a coating compound for these tests would have introduced a variable that may have affected results.

Mr. Bond's comment concerning the increase of resistance of bus connections fastened with steel bolts when compared to aluminum bolts after weather exposure is well taken. Difference of thermal expansion between the steel bolt and aluminum bolt may have been the major factor. The extreme degree of corrosion of the steel bolts, washers, and nuts (heavy scaling and partial disintegration) lead the authors to believe that part of the loosening may have been caused by loss of fastener material during about $1^1/_2$ years of weathering. The high rate of disintegration of the steel indicated that, if this was not the major factor in a short period of time, it would be the major factor in a longer period.

Although aluminum bolts are considered satisfactory for bus connection fastening they require the use of a bolt one size larger (1/2-inch steel bolt versus 5/8-inch aluminum bolt) and, to obtain sufficient contact pressure, the aluminum bolt must be tightened to approximately 80% of its ultimate strength. The additional space required for the aluminum fastener and the necessity of controlling tightening torque were the basis for conclusion 4 of the paper.

Predicting the Performance of a Wind Tunnel Regulating System Using an Analog Computer

R. HERBST
ASSOCIATE MEMBER AIEE

J. O. NICHOLS
ASSOCIATE MEMBER AIEE

F. W. KEAY
ASSOCIATE MEMBER AIEE

R. W. MILLER
ASSOCIATE MEMBER AIEE

THE USE of an analog computer, particularly one of the electronic differential analyzer type, in the analysis and design of feedback control systems has been well established in the past few years. Quite often, however, a detailed computer investigation of a control system is not necessary. For example, in some instances the control system may contain only a few components which can be represented by linear transfer functions, in which case the familiar methods of linear servomechanism analysis may be employed. It is also possible in some cases that it is just as economical to optimize a design on the test floor without benefit of a computer study.

In general, wind tunnel drive-regulating systems may not be similar in many respects, e.g., in ratings of the equipment involved and types of regulation required. Because of the large power requirements of a wind tunnel, the stability of the regulating system is of primary importance, and usually rapid response times can be sacrificed for stability improvement. Wind tunnels may have a wide operating range over which stability must be maintained. For example, it has been observed that a system may be stable through an acceleration cycle while under the control of current or power regulators, but may be unstable at operating speed. The control system usually contains components which have nonlinear characteristics such as saturation, clipping and backlash which cannot be neglected in the analysis. Therefore, linearizing a system and investigating the stability with respect to small excursions about some operation point cannot be regarded as safe procedure. The size of the equipment involved in a wind tunnel installation would, in general, prohibit the assembly of the complete system on the test floor. Also, the high costs of tests and equipment changes at the permanent installation site would not be desirable. For these reasons, a detailed analog computer study of a complex wind tunnel regulating system is extremely useful.

Paper 59-646, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19-21, 1959. Manuscript submitted August 18, 1958; made available for printing March 18, 1959.

R. HERBST, J. O. NICHOLS, F. W. KEAY, and R. W. MILLER are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

Computer Approach

The purpose of this paper is to discuss the analog computer approach in the analysis of a typical wind tunnel drive-regulating system. Before the system is ready for computer simulation, the basic design of the individual components is complete and their transfer characteristics have been determined. In the computer representation, the components are coupled in the designer's proposed arrangement. A typical starting cycle is then initiated and the system behavior is immediately available on time-scaled pen-recorded oscillograms. If a system proves to be unstable, the influence of various stabilizing measures which are suggested by a preliminary analysis employing approximate linear analytical methods (Bode or Nyquist diagrams) may then be studied. Since each component of the actual system has its computer counterpart, the parameters in the analog have physical identity, and the effect of variation of these parameters can readily be observed. It is important to bear in mind that the computer will not automatically optimize the performance or recommend the types of stabilization which may be required. The computer is simply a tool which is an aid to the engineer. It enhances the designer's intuitive "feel" for his system and refutes or confirms his predictions. Although this direct system simulation can be programmed for a high-speed digital computer, it is not as economical and convenient as the analog computer approach.

Regulating System

The wind tunnel under discussion consists essentially of a single wound-rotor induction motor which drives several stages of compressors. The total moment of inertia of the rotating machines and the load torque characteristic of the compressors are specified.

The regulating system will be discussed in a general manner with reference to the schematic diagram; see Fig. 1. The solid lines may be considered as flow lines of intelligence of the basic system, and the dashlines are feedbacks deemed necessary for stabilization.

The tachometer voltage which is proportional to induction motor speed is compared with a reference and the error is clipped by a diode circuit which is labeled "Clipper." The output of the clipper is then amplified by a preamplifier and a power amplifier, and this voltage in turn supplies the control windings of a shunt-tuned Rototrol. The Rototrol in turn controls the pilot motor which positions the liquid rheostat electrodes. The amount of rheostat resistance determines the total secondary resistance of the induction motor, and ultimately the torque and speed. In addition, the clipper signal

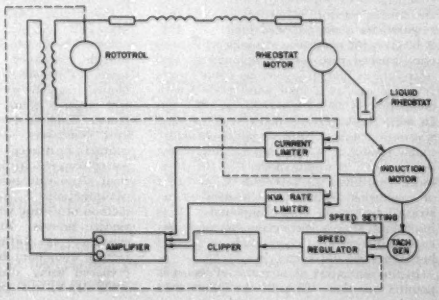


Fig. 1. Schematic of regulating system

can be annulled by the current limiter, which limits the secondary current of the induction motor, or by the kva (kilovolt-ampere) rate limiter which limits the rate of change of the induction motor primary current.

Three different types of regulators are associated with this sample system: speed, induction motor secondary current, and kva rate limiter. Some other types of regulators which may be present will be mentioned as examples. A dynamic braking current regulator could control the braking torque during a shutdown operation. In the case of a system containing more than one induction motor, current balance regulators may be required, i.e., one induction motor may be controlled by speed regulation and a second by current balance with the first. In some cases, a wind tunnel drive may be accelerated to operating speed by an induction motor, at which time the load is appropriated by a synchronous motor. If at any time the load were to exceed the synchronous motor rating, a kilowatt excess transfer regulator could be used to share the load with an induction motor. Appendix I gives the equations which describe the various components of the regulating system.

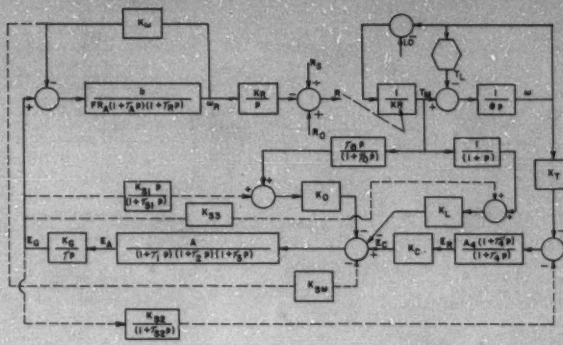


Fig. 2 (above). Block diagram of regulating system

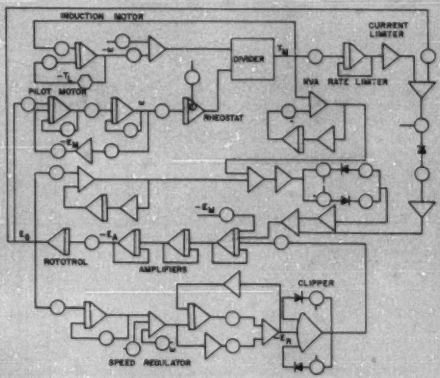


Fig. 3 (right). Computer analog of regulating system

Analysis and Results

Fig. 2 is a per-unitized block diagram constructed from the equations derived in Appendix I. The solid lines indicate components of the basic regulator, while dashed lines indicate components introduced as stabilizing measures. The analog representation shown in Fig. 3 follows from the block diagram. The initial gains and time constants of the various stabilizing feedbacks are selected on the basis of a preliminary analysis and previous experience. These parameters are then varied to determine their individual effects, and are then adjusted to obtain their optimum values.

Fig 4(A) is a pen recording of a typical acceleration cycle to rated speed without any stabilizing measures present. The system performs satisfactorily until the torque reaches the value of 1.25 per unit. At this time, the current limiter becomes active, the kva rate limiter becomes inactive, and a nonlinear instability appears simultaneously.

The preliminary linear analysis of the system revealed that the speed-regulating loop tended toward instability, and also that the two minor regulating loops involving the kva rate limit and the overcurrent limit were inherently unstable. This analysis also indicated the types of feedback that should be employed for stabilization. The stabilizing measures were introduced individually into the analog and the gains and time constants adjusted to their optimum value. Fig. 4(B) shows the effects of using a derivative feedback from the Rototrol internal voltage to the kva rate limiter. The system still has an unstable response but has improved in that rated speed is attained and the system passes through the "minor" loop instabilities. A direct feedback to the current limiter removed the minor loop instabilities; see Fig. 4(C). Finally a time delay network from the Rototrol internal voltage to the speed regulator proved effective in stabilizing the major loop; see Fig. 4(D).

A method suggested for obtaining the feedback signals from the back emf (electromotive force) of the pilot motor and the internal voltage of the Rototrol is shown in Fig. 5. For further discussion of Fig. 5, see Appendix II.

Conclusions

The use of an analog computer in conjunction with the analytical techniques of linear feedback control offers a powerful method for the investigation of systems of the type exemplified in this paper. The results of a study such as this do not dictate precisely the actual values of the feedback parameters which will appear in the final system. However, the performance curves of the final system, which in the case of a large wind tunnel would be obtained at the permanent installation site, should correspond very closely with the computer results. Because of probable inaccuracies in the physical data available for the computer study, minor "trimming" adjustments of some components will be necessary.

It is important throughout the computer study to bear in mind that only quantities which are physically obtainable

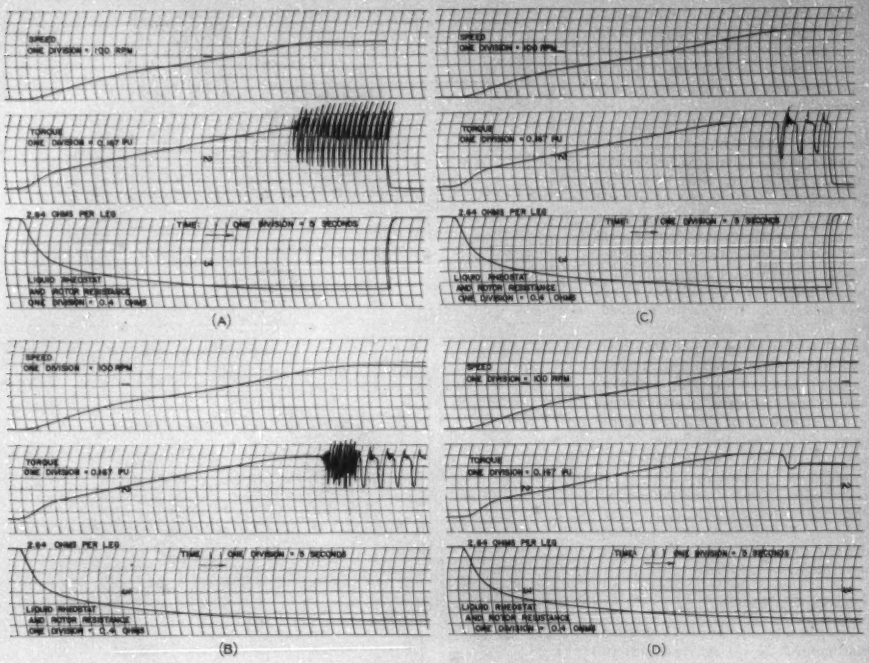


Fig. 4. Simulation of starting cycle

A—Without stabilizing measures
B—With feedback from internal Rototrol voltage to kva rate limiter
C—With feedback from internal Rototrol voltage to kva rate limiter and current limiter
D—With feedbacks from internal Rototrol voltage to kva rate limiter, current limiter, and speed regulator

should be used as feedbacks. Thus, a feedback from the internal voltage of the amplifier to the speed regulator should not be considered in the study even though this results in a favorable response, because the internal voltage of the amplifier is not available in the actual physical system (even though it is available on the computer analog).

Although the use of analytical techniques should be emphasized, the regulator discussed in this paper would be difficult to analyze completely by these methods, due to the nonlinear characteristic of the closed loop transfer function. However, with the aid of an analog computer, highly satisfactory results were obtained with relative ease.

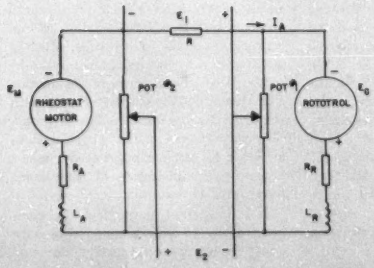


Fig. 5. Circuit for obtaining stabilizing feedbacks

Appendix I. Development of Equations

Induction Motor Equations

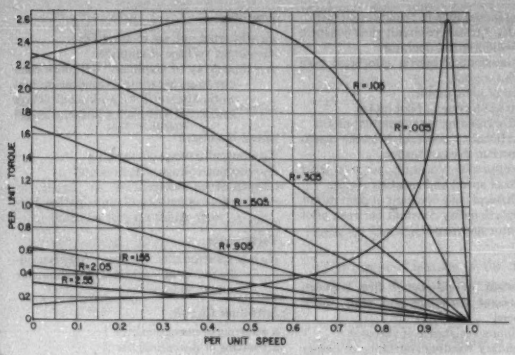
On a per unit basis

$$T_M = \frac{I^2 R}{S} \tag{1}$$

The torque of the induction motor is proportional to the secondary current if the secondary resistance is sufficiently large or if the speed of the motor is near synchronous speed; see Fig. 6(A). In this region of operation, equation 1 can be written as

$$S = KT_M R \tag{1A}$$

and



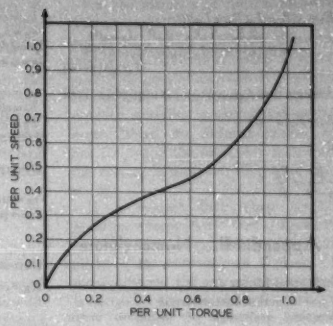


Fig. 6. A—Speed-torque curves of induction motor. B—Speed-load torque curve

$$\omega = 1 - S = 1 - KT_M R \tag{2}$$

The value of K can be determined from the torque-speed curves; see Fig. 6(A).

The induction motor torque-speed curves are calculated from the equivalent circuit shown in Fig. 7.

The equation of motion is

$$\Phi \frac{d\omega}{dt} = T_M - T_L \tag{3}$$

Rheostat Pilot Motor Equations

$$I_A = \frac{E_g - E_M}{1 + \tau_A p} \tag{4}$$

$$\lambda \frac{d\omega_R}{dt} = T_e - T_R$$

$$T_e = bI_A$$

$$T_R = F\omega_R$$

$$\omega_R = \frac{bI_A}{F(1 + \tau_R p)} \tag{5}$$

$$\omega_R = \frac{b(E_g - E_M)}{FR_A(1 + \tau_A p)(1 + \tau_R p)} \tag{6}$$

$$E_M = K_\omega \omega_R \tag{7}$$

Rheostat Equations

The relative position of the electrodes of the liquid rheostat is proportional to the angular displacement of the pilot motor shaft, and the resistance of the rheostat is proportional to the separation of the electrodes. In operational form

$$R = R_R + R_0 - \frac{K_R \omega_R}{p} \tag{8}$$

Rototrol Equations

Since the Rototrol has a self-excited shunt field, the internal voltage of the Rototrol is a function of control field current and shunt field current.

$$I_C = \frac{E_A}{R_C(1 + \tau p)}$$

$$I_S = \frac{E_g}{R_S(1 + \tau p)}$$

$$E_g = N_C K_g' I_C + N_S I_S K_g'$$

$$E_g = \frac{N_C K_g' E_A}{R_C(1 + \tau p)} + \frac{N_S K_g' E_g}{R_S(1 + \tau p)} \tag{9}$$

If the Rototrol has a shunt-tuned field,

$$N_S K_g' = R_S$$

$$E_g = \frac{N_C K_g' E_A}{R_C \tau p} \tag{10}$$

Amplifier

The amplifier has the following transfer function

$$E_A = \frac{AE_C}{(1 + \tau_1 p)} \tag{11}$$

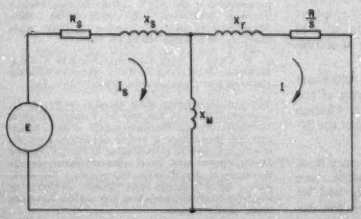


Fig. 7. Equivalent circuit of induction motor

Clipper and Limiter

Both the clipper and the limiter are nonlinear and hence do not lend themselves to a convenient analytical representation. However, their analog representation is quite accurate; see Fig. 3.

Speed Regulator

Incorporated in the speed regulator is an anticipation network; its transfer function is

$$E_R = A_r \frac{(1 + \tau_1' p)}{(1 + \tau_1 p)} (\text{reference} + E_T) \tag{12}$$

Tachometer

The tachometer device is linear throughout the region of operation and has the simple transfer function

$$E_T = K_T \omega$$

Appendix II. Stabilizing Feedback Circuit

The circuit shown in Fig. 5 is a modification of one proposed by Tustin.[1] It is intended as a means of obtaining voltages proportional to the back emf of the pilot motor and the internal voltage of the Rototrol. In the circuit of Fig. 5, it is assumed that negligible current will flow through potentiometers 1 and 2.

Applying Kirchoff's voltage law around the loop containing E_1 yields

$$\begin{aligned} E_1 &= E_M + I_A(R_A + L_A p) - \\ &\quad \alpha(E_M + I_A R + I_A R_A + I_A L_A p) \\ &= E_M(1 - \alpha) + I_A[R_A(1 - \alpha) - \alpha R] + \\ &\quad I_A L_A p(1 - \alpha) \end{aligned}$$

Assume that $I_A L_A p(1 - \alpha)$ is negligible; then

$$E_1 = (1 - \alpha)E_M$$

if

$$R = \frac{1 - \alpha}{\alpha} R_A$$

Solving for E_3 yields

$$E_2 = \alpha[E_g - I_A(R_R + L_R p)] - \beta(E_t - I_A R_R - I_A R - I_A L_R p) = E_g(\alpha - \beta) + I_A[R_R(\alpha - \beta) + \beta R] + I_A L_R p(\beta - \alpha)$$

Assume that $I_A L_R p(\beta - \alpha)$ is negligible; then

$$E_2 = E_g(\alpha - \beta)$$

if

$$\frac{\beta}{\alpha - \beta} = \frac{R_R}{R}$$

Appendix III. Nomenclature

Induction Motor

T_M = induction motor torque
R = total resistance of induction motor secondary circuit
I = secondary current of induction motor
S = slip
ω = speed of induction motor
$K^{1/2}$ = constant of proportionality between induction motor current and torque
Φ = total moment of inertia of wind tunnel drive
T_L = load torque of compressors; see Fig. 6(B)

Rheostat Pilot Motor

I_A = armature current
E_g = internal Rototrol voltage
E_M = motor back emf
R_A = tctal resistance in armature circuit
L_A = total inductance in armature circuit
$\tau_A = L_A/R_A$ = armature circuit time constant
$p = d/dt$ = differential operator
λ = total moment of inertia associated with pilot motor
ω_R = speed of pilot motor
T_e = electrical torque of pilot motor
T_R = load or resisting torque of pilot motor
b = proportionality constant between armature current and torque
F = proportionality constant between pilot motor speed and resisting torque
τ_R = mechanical time constant of pilot motor
$K\omega$ = proportionality constant between pilot motor speed and back emf

Rheostat

K_R = constant of proportionality between rheostat resistance and pilot motor speed
R_0 = maximum resistance of rheostat
R_R = secondary winding resistance of induction motor

Rototrol

E_g = internal voltage of Rototrol
E_A = control field voltage
R_C = control field resistance
τ = total time delay due to all field windings
I_S = shunt field current
I_C = control field current
R_S = shunt field resistance
N_C = number of turns on control field
N_S = number of turns on shunt field
K_g' = volts per ampere-turn gain of Rototrol

Amplifier

E_A = internal voltage of amplifier
E_C = input voltage of amplifier
τ_1 = time delay of amplifier. (If more than one stage is used in amplification network, there will be additional delays associated with each additional stage)

Speed Regulator

E_R = output voltage of regulator
E_T = output voltage of tachometer
ref = reference voltage
τ_4', τ_4 = time constants of anticipation network which may be adjusted

Tachometer

K_T = proportionality constant between induction motor speed and tachometer output voltage

Anti-hunt Circuit

α = setting of potentiometer 1
β = setting of potentiometer 2

References

1. Direct Current Machines for Control Systems (book), A. Tustin. The Macmillan Company, New York, N. Y., 1952, pp. 232–36.

2. Analog Computer Study of Wind Tunnel Drive, K. G. Black, R. J. Noorda. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, 1957 (Jan. 1958 section), pp. 745–50.

3. Electrical Transmission and Distribution Reference Book (book). Westinghouse Electric Corporation, Pittsburgh, Pa., 1950, pp. 191–92.

A New Chart Relating Open-Loop and Closed-Loop Frequency Responses of Linear Control Systems

C. F. CHEN
NONMEMBER AIEE

D. W. C. SHEN
MEMBER AIEE

CONTROL ENGINEERS generally adhere to the use of frequency response as the fundamental tool for feedback control systems. Through the use of the Nichols chart,[1] the open-loop characteristics can be quickly converted to the closed-loop characteristics. The performance of control systems can be estimated according to the gain and phase margins and the magnitude of peak M. Although actual performance may be quite different from that predicted by a certain value of a particular specification, yet by means of Meerov's method[2] it should not be difficult for engineers to determine the limits of applicability and to evaluate its effectiveness.

This paper attempts to develop a new chart which has several practical advantages over the well-known M-contours or Nichols chart. Some of the advantages are outlined as follows:

1. The graphical construction is easier, because it is based on the rectangular coordinates of both the open-loop frequency response, $G(j\omega)$, and the closed-loop frequency response, $M(j\omega)$. On the M-plane, the contours of constant real and constant imaginary parts of G are two families of orthogonal circles which intersect at a common point.

2. The G-locus when plotted on this chart gives the M-locus directly in terms of the rectangular or polar co-ordinates, whereas if the Nichols chart is used a replot of the M-locus is necessary.

3. To find the transient performance from the closed-loop frequency response, one usually resorts to the method devised by Floyd.[3] Floyd's method requires the determination of the real part of the closed-loop response, and this can be directly read from the new chart.

4. From this chart, the most common performance specifications of closed-loop systems, such as the gain margin, the phase margin, peak M, bandwidth, and rise time, etc., can be readily determined. In some cases, they are actually simpler than the use of conventional charts.

Construction of New Chart

In a unity feedback control system, such as shown in Fig. 1, the following relation obtains:

$$\frac{C}{R} = M\angle\alpha = \frac{|G|\angle\phi}{1 + |G|\angle\phi} \qquad (1)$$

Paper 59-647, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959. Manuscript submitted September 16, 1958; made available for printing March 18, 1959.

C. F. Chen is with Christian Brothers College, Memphis, Tenn., and D. W. C. Shen is with the University of Pennsylvania, Philadelphia, Pa.

The authors express their thanks to Dr. J. G. Brainerd, Dr. Y. H. Ku, and Dr. Lambert Thomas for their constant encouragement.

Note: Attention has been directed to the possibility that the charts, and uses of them similar to those described in the paper, may have appeared earlier in the literature of the Russian *Automatika i Telemekhanika*. In *Automation and Remote Control*, the available English translation, the authors have not found similar charts.

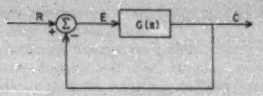


Fig. 1. Typical configuration of feedback control system

In terms of rectangular co-ordinates, $G/\Phi = a+jb, M/\alpha = U+jV$, equation 1 can be resolved into two equations involving the real and the imaginary parts respectively, i.e.,

$$U = \frac{a^2+b^2+a}{(1+a)^2+b^2} \tag{2}$$

$$V = \frac{b}{(1+a)^2+b^2} \tag{3}$$

Elimination of b from equations 2 and 3 gives

$$\left(U-\frac{2a+1}{2(1+a)}\right)^2+V^2=\left(\frac{1}{2(1+a)}\right)^2 \tag{4}$$

and elimination of a gives

$$(1-U)^2+\left(V-\frac{1}{2b}\right)^2=\left(\frac{1}{2b}\right)^2 \tag{5}$$

Equation 4 represents a family of constant-a circles having centers on the U-axis at $2a+1/[2(1+a)]$ and radii of $1/[2(1+a)]$. Equation 5 is that of constant-b circles having centers on the $U=1$ line at $1/2b$ and radii of $1/2b$. These two families of circles are orthogonal and they intersect at the common point $U=1$, $V=0$.

The chart is shown in Fig. 2. Contours of a and b are plotted in solid lines, while the polar co-ordinates of M/α are drawn in dashed lines. Symmetry exists in the plane containing the a and b contours with respect to the U-axis or $\alpha=0$, a 180-degree line only. The common point of intersection is the limit of a and b contours when either a or b approaches infinity. Physically it means when the open-loop gain G approaches infinity, the closed-loop gain M approaches unity. The b contour degenerates into a straight-line which is the U-axis, when b approaches zero. As mentioned previously, this line is the line of symmetry for the b contours. Physically, if the open-loop response G is a pure gain with zero phase angle, so is the closed-loop response M. On the other hand, the a contour when a approaches zero is a circle having center at $U=1/2, V=0$, and with a radius of one-half unit. All a contours for positive values of a, i.e., $0 \leqslant a \leqslant \infty$, are confined within the $a=0$ contour. For $-1<a \leqslant 0$, the a contours expand to the left-hand side until $a=-1$, where it degenerates into a straight line which is $U=1$. For $-\infty<a<-1$, a contours are shrinking circles on the right-hand side of $U=1$. The limit is again the point $U=1$, $V=0$, as a approaches $-\infty$.

It is interesting to note that the region within the contour $a=0$ is a replica of the so-called Smith chart[4,6] well known in high-frequency transmission-line theory.

Stability Considerations

The essential feature of the new chart is that the closed-loop response manifests itself directly in terms of either the rectangular or the polar co-ordinates, once the open-loop response is plotted on the a and b contours.

Leonhard's method[6,7] may be used to determine the stability of control system in the M-plane. For the case of the open-loop transfer function having no poles in the right-hand half-plane, his criterion is as follows: If the M-locus in going from zero frequency toward infinity, sweeps through as many quadrants (in a negative direction) as the highest power of s (the Laplace transform variable) in the denominator of M diminished by the highest power of s in the numerator, the control is stable.

For the majority of control problems that are stable when the control system loop is open, Bode's network theorems[3] provide the basic information necessary to simplify greatly the work of a designer in obtaining a stable system. A new stability criterion is established as follows: When the M-locus crosses the line $U=1/2$, if the value of V is positive, the system is unstable; otherwise, it is stable.

It is evident that when $U=1/2$, equation 2 becomes $a^2+b^2=1$; hence the line $U=1/2$ corresponds to the unit circle in the conventional G-plane. In many systems, the sign of a at crossover is negative; the sign of b follows that of V. Therefore the foregoing criterion is equivalent to the familiar statement that "systems having positive phase margin when their transfer function crosses the unit circle are stable, whereas systems with negative phase margin when their transfer function crosses the unit circle are unstable."[9]

It follows that phase margin is obtained from the new chart by measuring $\tan^{-1} b/a$ at the crossover. The gain margin is given by the reciprocal value of a at the point where the M-locus crosses U-axis, since the U-axis coincides with the $b=0$ contour.

The advantages of the new chart are

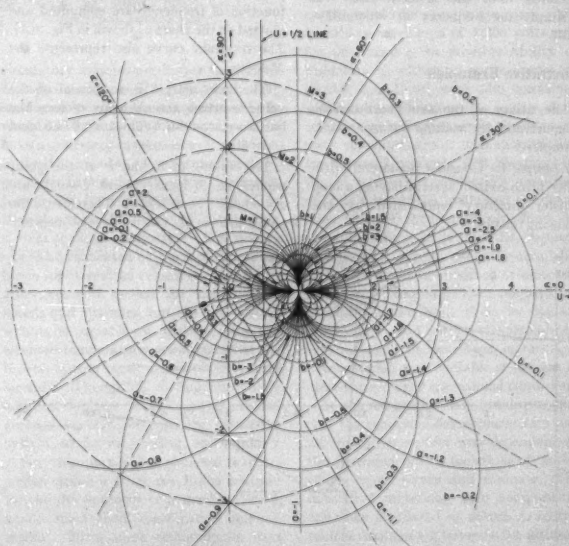


Fig. 2. The new chart

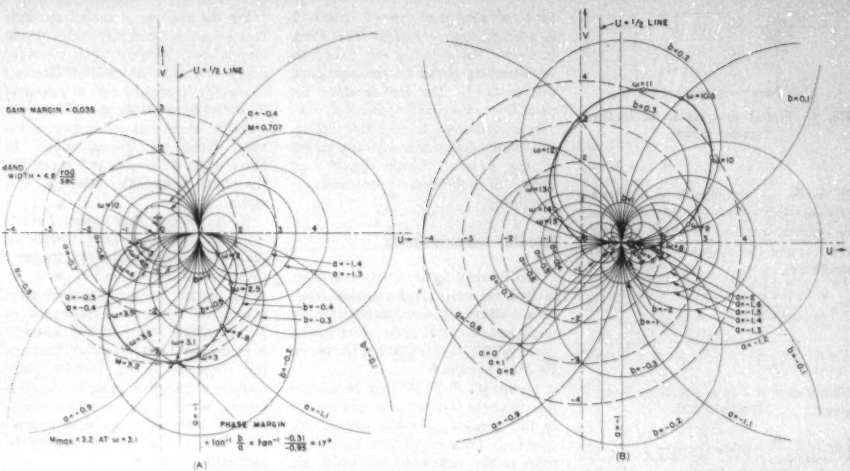


Fig. 3. Frequency response. A—Of stable system. B—Of unstable system

also obvious when the quantity of peak M is to be evaluated. It can be accurately read from the M-locus itself. Likewise, bandwidth can also be determined from the M-locus itself. It is simply the frequency at which $M = 0.707$.

Illustrative Examples

The utility of the new chart can be demonstrated by working out two simple examples.

Example 1: Consider an aircraft propeller pitch-control system having a loop transfer function represented by the following relation:

$$G = \frac{k}{s(s+1)(1.2\times10^{-4}s^2+3\times10^{-3}s+1)}$$

Assume $k = 10$, the real and the imaginary parts of the foregoing equation as a function of frequency are computed and plotted on the chart as shown in Fig. 3(A). The resultant curve also represents the M-locus.

The gain margin is reciprocal of the value a when the M-locus crosses the line $V = 0$ or $b = 0$. Thus, $a = 0.035$, and $1/a = 30$.

The values of a and b at the point where the M-locus crosses $U = 1/2$ are $a = -0.95$, $b = -0.31$. Therefore, the phase margin is $\tan^{-1} b/a = 17$ degrees.

Since the sign of b is negative, according to the criterion just given, the system is stable.

The peak value of M and the bandwidth are directly read from the chart to be 3.2 and 4.8 radians per second respectively.

To find the transient response, evaluate the following real form of the Fourier integral:

$$h(t) = \frac{2}{\pi}\int_0^\infty R(\omega)\cos\omega t\, d\omega \qquad (6)$$

which is the basis of Floyd's graphical method. In equation 6 $R(\omega)$ represents the real part of the closed-loop frequency response $M(\omega)$.

The function $R(\omega)$ is directly read from the chart and is plotted in Fig. 4. The corresponding time response to a unit step input is shown in Fig. 5.

It should be noted that when an M,α chart is used, an auxiliary chart of con-

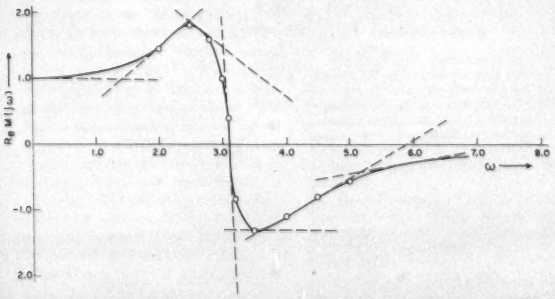


Fig. 4. Real parts of frequency response of example 1

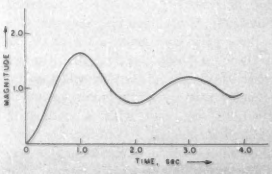


Fig. 5. Transient response of example 1

tours is necessary, as suggested by Floyd, in order to obtain $R(\omega)$.

Although this system is stable, the low phase margin and low damping ratio indicate that the system is rubbery with excessive overshoot and long settling time.

Example 2: A system with

$$G=\frac{40}{S(25\times10^{-3}S+1)(625\times10^{-6}S+1)}$$

is plotted on the chart shown in Fig. 3(B). The M-locus crosses $U=1/2$ with positive value of b, and therefore the system is unstable.

Conclusions

A new chart is developed which correlates the open-loop and closed-loop responses of a unity feedback control system.

Graphically, M,α contours look like the old ultrahigh-frequency transmission line chart and the new chart looks like the Smith chart if the latter is extended to the region of negative real parts.

Aside from some of the practical advantages of the new chart already mentioned, it is hoped that this paper may broaden the viewpoint of engineers and students alike when they make a study of the conventional M,α contours or the inverse complex plane plot as recommended by Marcy[10] and Whiteley.[11]

References

1. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, Inc., New York, N. Y., 1947, pp. 158–62, 179–86.

2. Estimating the Performance of Control Systems According to Their Stability Margin with Respect to Modulus, Phase and the Quantity M (in English), M. V. Meerov. *Automatika i Telemekhanika*, Moscow, U.S.S.R., vol. 17, no. 10, 1957, pp. 965–69.

3. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley & Sons, Inc., New York, N. Y., 1948, pp. 339–46.

4. Transmission Line Calculator, P. H. Smith. *Electronics*, New York, N. Y., vol. 12, Jan. 1939, pp. 29–31.

5. An Improved Transmission Line Calculator, P. H. Smith. *Electronics*, vol. 17, Jan. 1944, pp. 130–33.

6. Die Selbsttätige Regelung (book), A. Leonhard. Springer-Verlag, Berlin, Germany, 1957, pp. 233–47.

7. Neues Verfahren zur Stabilitätsuntersuchung, A. Leonhard. *Archiv für Electrotechnik*, Springer-Verlag, vol. 38, 1944, pp. 17–28.

8. Network Analysis and Feedback Amplifier Design (book), H. W. Bode. D. Van Nostrand Company, Inc., New York, N. Y., 1945, pp. 312–19.

9. Servomechanism and Regulating System Design (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., 1951, pp. 292–97.

10. Parallel Circuits in Servomechanisms, H. Tyler Marcy. *AIEE Transactions*, vol. 65, Aug.-Sept. 1946, pp. 521–29.

11. Theory of Servo Systems, with Particular Reference to Stabilization, A. L. Whiteley. *Journal*, Institution of Electrical Engineers, London, England, vol. 93, pt. 2, 1946, pp. 353–67.

12. A Quick Method for Estimating Closed-Loop Poles of Control Systems, K. Chen. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, May 1957, pp. 80–87.

Generator Insulation Systems Development for Hypersonic Aircraft

W. B. PENN, MEMBER AIEE — R. L. BALKE, MEMBER AIEE — F. M. PRECOPIO, ASSOCIATE MEMBER AIEE

PRESENT and future aircraft requirements for electric power are constantly increasing, with operational requirements of speed, altitude, and equipment imposing greater loads on the electric generating systems. The need therefore exists for generators capable of operating at higher temperatures than those presently available.

Various methods have been considered and are being used to extend the operational range of present insulations in the electric power generators by techniques for extracting heat. However, these impose a weight penalty upon the system in terms of the weight of cooling fluid and circulatory systems needed, or the weight and wind resistance of the heat exchanger employed by their cooling devices.[1,2] It would be desirable and advantageous to have a generator insulation system which in itself is capable of withstanding temperatures in the vicinity of 500 C (degrees centigrade) for continuous operation, so that the excessive weight of any cooling system could be eliminated. This system would be ventilated by the ram air available in flight just as corresponding generators are now cooled for conventional speeds and altitudes, but the new system would be capable of withstanding the elevated temperatures encountered at the higher speeds and altitudes of future aircraft. Though oil cooling was finally adopted for this machine, the original goal of a high-temperature insulation system, as for air cooling, was retained.

For this purpose, a proposal was accepted under a U. S. Air Force contract for the development of a machine which would meet their specification requirements. Since these requirements were presented in terms of performance, i.e., output under defined conditions of altitude, air inlet temperature, and load, it was necessary to tie in the best estimates of design criteria to permit an estimate of the temperatures which the windings would attain under these conditions. These temperature estimates formed the basis of a test cycle which the insulation system would be required to meet. Dielectric requirements of 1,200 volts rms for one minute, to measure ability of insulation to withstand various environmental conditions, were also contained within the specification.

The technical approach in the development of components of such an insulation system was 3-pronged: first, the capabilities of presently commercially available materials would be evaluated to determine whether they could withstand the simulated duty cycle for the required service life; second, presently available materials would be modified or adapted, if necessary, to extend their temperature capabilities to permit them to operate successfully; and third, a completely new system would be developed in terms of new materials not available at present to meet the required operational conditions.

It was recognized early in this program that the successful development of a useful insulation system was dependent on the development of a technique of testing which would screen and evaluate, first, individual materials, then components, and then simulated or actual structures such as would be contained in the ultimate design of the machine. For this reason procedures were considered and used

Paper 59-864, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 23, 1959.

W. B. Penn, R. L. Balke, and F. M. Precopio are with the General Electric Company, Erie, Pa.

The authors wish to acknowledge the technical assistance and the benefit of consultation provided by J. H. Terry, J. A. Dimond, F. E. Johnson, P. A. Rizzo, and J. T. Bateman, all of the General Electric Company.

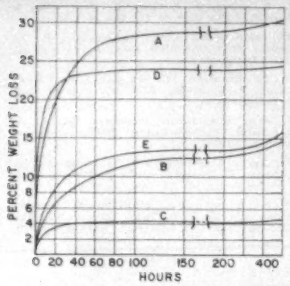


Fig. 1. Weight loss of conventional silicone resins at 300 C as function of time

based on previously successful work in the evaluation of aircraft generator insulation systems[3] as well as procedures developed and incorporated in AIEE Standards 510 and 511.[4,5] The test duty cycles of the contractor's specification were translated into time and temperature co-ordinates from which a test cycle for laboratory evaluation was evolved. To permit closer measurement of the deterioration process, moisture was included in the cycle of time at temperature, vibration, and high-potential testing so that a procedure analogous to AIEE Standard 510 was developed.

So that actual components could be evaluated, a test vehicle was developed which could be assembled by conventional manufacturing techniques and operators, and which permitted a number of measurements to be made. This fixture evaluated factory ease of handling of the components, assembly techniques, and treating procedures, while measuring the ability of the insulation system to withstand the rigorous test cycle established. With this general approach established, the details of the development program proceeded as described in the following.

Development of Materials

The mechanical and electrical properties of organic or metal-organic insulation materials are affected by such environmental conditions as humidity, vibration, thermal shock, corona, etc., but the most important deteriorating factor, either alone or in combination with these, is high temperature. As operating temperatures are raised, the number of materials which can function properly for reasonable periods diminishes rapidly. At 100 C, there are scores of organic polymers which may be used to give long life as electrical insulators under severe mechanical and chemical conditions. Treated natural fibers, polyamides, polyesters, polymers of many vinyl compounds, epoxides and phenolics are representative of these. At 200 C, this list has narrowed to no more than three or four polymer types (phenolics, special epoxides, silicones, polytetrafluoroethylene); and above 300 C, the silicones alone are serviceable but only if properly formulated and if the length of service required is very short. Therefore, as the use temperature exceeds 300 C, it is necessary to turn to inorganic or ceramic materials for electrical insulation. Their mechanical properties leave much to be desired, the electrical properties fall off badly at very high temperatures, and processing presents many difficulties. However, they inherently possess high thermal stability and therefore present a promising approach to the solution of the ultrahigh-temperature insulation problem.

Silicone Insulations

Because of the afore-mentioned disadvantages of ceramics to be overcome, and since more conventional methods and materials were not to be overlooked, two insulation materials programs were carried along in parallel with the ceramic developments.

The first was an evaluation of the top limit to be expected from conventional silicone varnishes. Motorette tests have been conducted for some time in connection with the development of insulation systems for aircraft generators.[3] The most promising high-temperature systems have been based on silicones, mica, and glass. The results of these tests indicated that a possibility existed that a silicone system could meet the requirements of this program if a silicone varnish of truly superior thermal resistance could be found. Five commercially available silicone resins were chosen for evaluation, and weight-loss tests were conducted at 300 and 450 C. Fig. 1 shows the weight-loss results at 300 C and it is evident that silicone resins can differ greatly in thermal stability. An attempt to heat-age the resins at 450 C gave very bad degradation in one hour (see Fig. 2) and the tests were not continued further. The most stable varnish was then used to treat motorettes constructed with glass-served silicone-treated magnet wire and glass-mica-glass ground and phase insulation. After baking 8 hours at 200 C, the motorettes were subjected to 6 hours at 275 C, 6 hours at 350 C, and 24 hours at 500 C. The varnish remained intact through the 350 C run but degraded to a sandy deposit after the 500 C exposure.

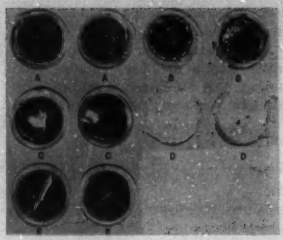


Fig. 2. Photograph of conventional silicone resins after 1 hour at 450 C

Leakage-current measurements at 1,200 volts to ground showed that the systems were electrically sound after the 350 C aging but high leakage (30 microamperes) was observed after the 500 C aging. It was concluded that conventional silicone systems could not be relied upon for use at 500 C but that, for limited exposure times, use up to the vicinity of 400 C might be feasible.

Filled Solventless Silicones

The second approach was based on an evaluation of filled solventless silicones reported to be capable of continuous operation at well over 300 C. Two different solventless silicones were compounded with such fillers as alumina, silica, barium silicate, zirconium silicate, zirconia, mica flakes, glass shorts, and glass beads, to obtain the highest loadings which would still be pourable. Air inhibition of cure was observed and it was necessary to cure the materials in nitrogen. A more serious problem was the tendency of the fillers to settle out during cure, leaving a layer of clear resin which cracked extensively due to thermal shock.

Table I. Specific Resistivity and Dielectric Strength of Three Ceramic Glazes on Steel as a Function of Temperature

Temperature, C	Type of Glaze Designation A	B	C
Specific Resistivity, Ohm-Centimeters			
RT*	2.8×10^{11}	5.6×10^{9}	1.6×10^{13}
100	6.6×10^{10}	2.6×10^{9}	1.3×10^{12}
200	3.1×10^{9}	5.1×10^{7}	2.8×10^{9}
300	7.0×10^{8}	3.4×10^{7}	1.4×10^{9}
400	3.8×10^{8}	2.5×10^{7}	5.0×10^{8}
500	3.2×10^{7}	5.5×10^{6}	1.2×10^{8}
600	2.6×10^{6}	4.2×10^{6}	2.8×10^{8}
Dielectric Strength, Volts per Mil			
RT*	570	260	290
300	150	110	120
500	110	60	60
600	90	40	60

* Room temperature.

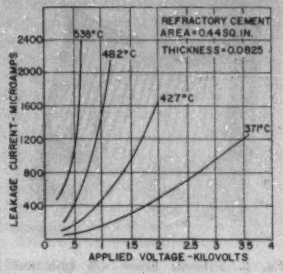


Fig. 3. Leakage current as function of voltage and temperature for typical low-porosity cement

This problem was overcome by the use of small percentages of a thermally stable suspending agent. The materials were cast into 2-inch-diameter 1/4-inch-thick disks which were then heat-aged at 300, 500, and 600 C. The resistance of the resin-filler combination to thermal degradation varied greatly. A few samples disintegrated after a short time at 300 C and some showed unexplainable strength after aging at 600 C, where presumably an interaction has taken place. Results of statorette testing of some of these materials will be discussed later.

Ceramic Insulations

The major effort in this program was placed on the development of ceramic encapsulating compounds and glazes. The glazes were studied as sealing materials for the encapsulating cements, and stator structures, and as bonding agents for the glass-mica sheet insulations and magnet wire. A large number of vitreous glazes were screened by resistivity versus temperature measurements, dielectric breakdown, humidity corrosion tests, salt spray, compatibility with substrate materials, thermal shock, and thermal aging. Of the more than 60 glazes tested, 3 were found to withstand 100 hours in 100% RH (relative humidity) at 50 C, 100 hours in salt spray, thermal shock from 600 C to room temperature, and heat aging for 100 hours at 600 C without any signs of spalling, pitting, loss of gloss, or other evidence of failure. Specific resistivity and dielectric strength of these materials as a function of temperature were measured by coating them on steel panels and attaching a 2-inch-diameter electrode to the glaze over a coat of conducting paint. The results are shown in Table I.

The use of refractory cements for impregnation of windings to give a solid void-free structure would be very desirable. The cement could be applied in a wet-winding procedure or by vacuum-pressure cycling techniques. Such cements have been limited by high porosity, poor electrical properties, and short pot life.

This work has been directed at the development of low-porosity cements with reasonable pot life and acceptable electrical properties. It was found that combinations of refractories, binders and frits were capable of giving cements of very low porosity. This was accomplished by firing the cement above the melting point of the glass frit. Cements have been formulated that show less than 1% moisture absorption after boiling in water for 2 hours. The pot life of these cements can be varied from minutes to days. Leakage current as a function of voltage and temperature of a typical low-porosity cement is given in Fig. 3.

Micaceous and Quartz Insulations

In addition to these materials programs, work is in progress on the fabrication of solt liners made from mica paper saturated with a borate-phosphate complex. Slot liners have been molded successfully and used in making statorettes and test assemblies. Stability and retention of adequate dielectric properties have been demonstrated up to 600 C. Attempts are also being made to develop a mica mat slurry similar in composition to the slot liner material for use as an inorganic treating varnish. Moisture resistance is the main difficulty. A flexible flake mica composite bonded with an inorganic resin has successfully demonstrated its ability to withstand 500 C and retain moisture resistance.

The recent commercial availability of quartz fibers has led to a consideration of its use as serving for magnet wire. The build approximates that obtained with 150 series glass filaments. Test results have not been obtained with this magnet wire at this time.

Fig. 4. Photograph of statorette fixture and coils, prior to assembly

Development of Systems and System Tests

The evaluation of materials alone or in simple combinations is useful and important for screening and selecting the components to comprise an insulation system. More often, however, the manner in which materials are combined and the processing necessary to achieve a system will cause materials to perform differently in a system from the way they do as elementary components. Therefore, the need was recognized for a system test that would provide a means to study processes for combining materials and a means to measure system performance.

Development of Test Fixture

To facilitate evaluation of the insulation systems developed for this generator, a test fixture was developed embodying the electromagnetic features and configuration of the anticipated generator-stator assembly. This was done primarily to introduce the element of realism in component fabrication and placement, material processing, and finally measuring performance of the combined system under the stringent environmental conditions to be imposed on the generator. For instance, use of a realistic test fixture provides experience in handling some of the more unusual materials selected to meet the generator requirements and

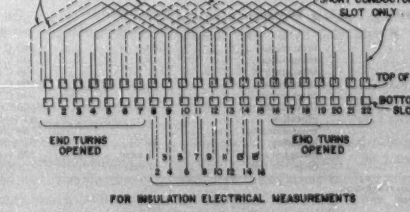


Fig. 5. Schematic diagram of statorette winding

permits study of best processing for coil bonding and environmental protection of the windings. In addition, a fixture closely resembling the generator stator provides a method to study design parameters of insulation thicknesses and spacing considered critical at the specified high altitudes and temperatures. Although test fixtures were also developed for other smaller assemblies of the generator containing insulations, this paper concerns only the stator assembly since it contains the majority of insulations, most of which are common to other assemblies.

A stator section of a production aircraft generator best met these conditions and was selected as the main element of the test device. Therefore, a production stator was selected which was approximately 10 inches in diameter, with a bore of 8.00 inches, stack length of 4.25 inches, and 66 skewed rectangular open slots having dimensions of approximately 0.190 inch wide by 0.406 inch deep. One-third sections of this production stator, simulating the anticipated high performance generator stator design, were mounted on stainless steel racks to facilitate handling and to provide a frame around the stator winding conforming closely to the generator design.

A view of this portion of the test fixture is shown in Fig. 4. It provided 22 stator slots and, by utilizing the same preformed hairpin shape coil for lap-type double layer winding, and the same span and forming fixtures employed in the selected production stator, two full coils with two conductors per slot could be inserted into these slots. A partial winding from a third coil was necessary to complete the winding. However, in so doing, a number of slots contained one conductor and a length of conductor shorter than the stack length was inserted into these slots to achieve compactness. The test fixture was designed to permit a number of electrical measurements by opening the involute ends of all coils on one end of the winding. For example, with the described configuration it is possible to obtain 8 individual measurements of conductor-to-conductor in the slots, 15 measurements of conductor-to-stator iron, and a multitude of crossover measurements in the end-turn windings. Fig. 5 illustrates this condition by schematic diagram. The developed test fixture was called a "statorette"; see Fig. 6.

The study of insulation thicknesses and spacing required can be made with this fixture by varying the conductor cross-section area. At the offset of this program, a 0.145- by 0.145-inch conductor size was selected with a 0.012-inch insulation build. Initial slot tube insulation thicknesses were 0.010 inch nominal. By allowing 0.003 inch for lamination stagger and skew, a reasonably tight slot was obtained. Increase or decrease of insulation thicknesses, which may be required for study purposes, can be achieved by a change in conductor cross section without affecting the desired compactness of slot.

To summarize, the function of the statorette is considered to be related to that of the motorette, now adopted for evaluation of insulation systems by the test procedure in the AIEE Standard.[4] The statorette is intended to satisfy the more specific requirements of this particular aircraft generator stator design.

Test Cycle

A test cycle was established primarily on the basis of the flight profile and endurance time of the high-performance aircraft. Conditions of test were patterned after those imposed on the generator with consideration for winding temperature rise, operating voltages, and conditions to provide best evaluation of the insulations.

Temperature, thermal shock, humidity, vibration, and the combination altitude and temperature were considered the most significant and critical of the many environmental conditions imposed on the generator. It was considered more practical to treat the combination altitude and temperature exposure as an independent test not to be included in the proposed cycle since such an exposure presents major evaluation problems in itself. Therefore, a tentative test cycle for evaluation of insulation systems by statorette test was established as shown in Table II.

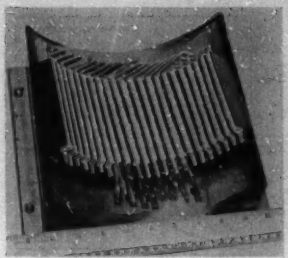


Fig. 6. Statorette fixture with coils and insulation inserted and with temporary hold-down slot wedges of Teflon

Steps 2 through 11 of the cycle just described are repeated 10 times. This is representative of the endurance cycle and the test required of the generator. To satisfy a test condition for the overload requirement of the generator, an exposure to 600 C for 30 minutes is included after alternate 500 C exposure conditions. A test has not yet been developed for the altitude-temperature condition.

Preliminary Results

Three statorettes have been built and have completed four of the required ten cycles of heat aging, vibration, and humidity. Leakage current measurements are being made at the various stages of each test cycle to provide a nondestructive test tool for following possible degradation of the systems.

The procedure appears quite capable of differentiating between various systems, as shown by leakage current measurements taken at 1,200 volts in the fourth cycle of test. The room temperature readings were taken after a dry-out period which followed the exposure to moisture of the third cycle. The readings at 500 C

Table II. Tentative Test Cycle

Sequence	Exposure	Time, Hours	Condition
1	temperature	300	200 C oven
2	temperature	8	300 C oven
3	temperature	1	300 C to 500 C
4	temperature	2	500 C oven
5	temperature	1	500 C to 300 C
6	thermal shock		300 C to 25 C
7	vibration	1	60-cycle 30-mil amplitude
8	measurements		IR* and leakage current
9	humidity	48	95% RH, 95 F
10	dielectric test	1 min†	208 volts, wet
11	dielectric test	1 min	1,200 volts, after drying

* Current resistance.
† Minute.

Table III. D-C Microampere Leakage on Three Statorette Insulation Systems Measured at 1,200 Volts

System	Room Temperature	500 C	95% RH	
A	0	10–25	>2,500	
B	0	5–10	70–90	p-to-g*
			400–450	p-to-p†
C	0	3–5	8–12	p-to-g
			4–10	p-to-p

* Phase-to-ground.
† Phase-to-phase.

followed the fourth exposure of 8 hours at 300 C plus 2 hours at 500 C. The final readings were taken in the humidity cabinet after vibration and exposure for 48 hours to 95% RH at 100 degrees Fahrenheit. Leakage is given in microamperes at 1,200 volts, where this voltage could be attained. On one sample, excessive leakage prevented reaching this voltage. Readings were taken phase to phase, and phase to ground; see Table III. These readings have been maintained throughout the test thus far.

Summary

This presentation is essentially a status or progress report. The detailed test results will be reported at the conclusion of the program. To summarize the work to date:

1. A test vehicle has been developed which simulates an actual generator structure to permit evaluation of factory handling properties as well as materials. It permits observation and measurement of the ability of the insulation system to withstand the deteriorating effects of a testing cycle based on specification conditions and requirements of temperature, vibration, humidity, and high-potential test.

2. There are available and in process materials and component structures whose adequacy is now being determined, and whose properties can be modified, if necessary or desirable, in order to produce a satisfactory insulation system, meeting the requirements of the generator specification for supplying electric power to hypersonic aircraft.

References

1. Ram-Air Cooling Systems for Aircraft Generators. R. M. Moroney. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, Sept. 1957, pp. 217–21.

2. Thermal Consideration of Generators in High-Speed Aircraft. A. Kusko. *Ibid.*, pp. 205–08.

3. The Motorette as a Tool for the Development of Improved Insulation Systems. W. B. Penn. *Ibid.*, pt. III-B (*Power Apparatus and Systems*), vol. 73, Dec. 1954, pp. 1505–07.

4. Test Procedure for Evaluation of Systems of Insulating Materials for Random-Wound Electric Machinery. *AIEE Standard No. 510*, Nov. 1956.

5. Test Procedure for Evaluation of Systems of Insulating Materials for Electric Machinery Employing Form-Wound Pre-Insulated Coils. *AIEE Standard No. 511*, Oct. 1956.

Battery Impedance: Farads, Milliohms, Microhenrys

E. WILLIHNGANZ
NONMEMBER AIEE

PETER ROHNER
NONMEMBER AIEE

THE ELECTRICAL RESISTANCE of storage batteries is important to two groups of people. Designers of telephone equipment need to know the resistance to alternating current in order to evaluate the cross talk between simultaneous telephone conversations which must pass through the central station battery. They must also be able to evaluate the amount of filtering required to remove hum, noise, and switching pulses from the output of the charging and switching equipment. In addition, designers of central power stations need to know the resistance of the battery which must supply short pulses of power to operate electric switches.

Surprisingly little has been published about the electrical resistance of batteries. G. W. Vinal[1] reviews what has been published. Most of the published material deals with methods of test and the cause of discrepancies between the various methods. There is, however, practically no data of interest to the engineer dealing with commercial types of batteries.

Paper 59-823, recommended by the AIEE Chemical Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting, and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 23, 1959; made available for printing May 15, 1959.

E. Willihnganz and Peter Rohner are with C & D Batteries, Inc., Conshohocken, Pa.

This work, therefore, has been undertaken to supply actual measured resistance values on batteries which are of engineering interest. A series of telephone-type lead-calcium batteries was selected for this test since they appear to be the type of greatest importance today.

Definition of Terms

The battery properties to be discussed are illustrated in Fig. 1. This is an oscillogram of the cell voltage during a 1-second discharge at about the 1-hour rate. When the discharge begins there is a sudden decrease of voltage because of the electrical resistance, and then a gradual decrease in voltage caused by chemical changes occurring at the surface of the active materials. When the discharge is interrupted, the voltage recovers in the same sequence. The electrical resistance can be calculated from either the sudden voltage change at the beginning of the discharge, or the sudden change at the end of discharge.

The rest of the curve represents the result of chemical changes. The chemical changes within the battery produce a counter emf (electromotive force) called polarization. This polarization cannot be precisely duplicated by a combination of the ordinary circuit elements, but it can be approximated by a capacitance shunted by a leakage resistance. The capacitance is responsible for the initial slope of polarization curve, while the leakage resistance is responsible for the curvature of this line.

The battery also possesses a measurable inductance. The initial part of the polarization curve shows a slight irregularity. Where the current is large, or the pulse length short, this irregularity can be clearly identified as the result of battery inductance. When large batteries are being studied, the effect of inductance assumes a major importance.

The equivalent circuit of a battery therefore contains a resistance, an inductance, and a capacitance, as shown in Fig. 2. This capacitance is shunted by a nonlinear resistor. The battery resistance has been divided between the metal resistance R_m and the acid resistance R_a. The resistance of the active material has been ignored, since it is believed to be negligible.

Because of the wide range of battery sizes tested, results were reduced to an amp (ampere)-hour basis. The measured resistance was multiplied by the ampere-hour rating of the battery, and the measured capacitance was divided by the ampere-hour rating of the battery. The inductance measurements, however, are tabulated directly without calculation because they varied only slightly with battery size.

Bridge

The oscillograms were used principally to define the terms used. All batteries were first tested using oscillograms, but this report contains principally Wheat-

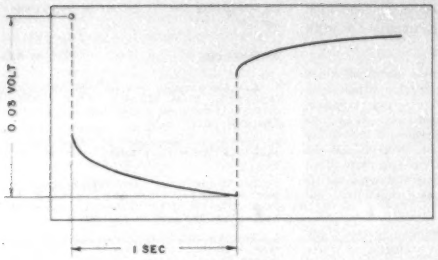


Fig. 1. Battery voltage during square wave pulse

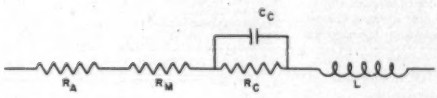


Fig. 2. Equivalent circuit of a battery

stone Bridge measurements, because they are not only more precise, but also considerably easier to obtain.

The bridge is shown in Fig. 3. The test cell was bolted directly to an ammeter shunt which served as a standard resistor R_S. The shunt was then bolted to a straight copper rod of negligible resistance which served as a variable inductor, L_2, by using a travelling contact. The ratio arm, R_4, was a 100,000-ohm fixed resistor. The measuring arm consisted of an adjustable resistor, R_3, zero to 1 megohm, and an adjustable capacitor 0 to 1 microfarad, C_3. The null indicator is a cathode-ray oscillograph with vertical plates across the bridge and horizontal plates across the variable frequency oscillator.

The bridge was used only up to 5,000 cycles because of battery properties. The bridge was checked by comparing ammeter shunts and was found to be within 1% over a frequency range up to at least 50,000 cycles.

For measurements, the bridge was first balanced for resistance and inductance at 5,000 cycles, with the capacitance omitted. The inductance setting was then left unchanged while the bridge was balanced for resistance and capacitance at 1,000 cycles.

Inductance

The inductance of batteries is summarized in Fig. 4 and is about 0.1 microhenry. The 10 to 1,000-amp-hour part of the chart describes batteries having single posts. The 300 to 1,600-amp-hour section describes batteries with double posts. For these larger batteries, a heavy copper bus-bar was firmly bolted to both positive posts, and another similar bar bolted to the negative posts. Bridge connections were made to the mid-point of these copper bars. The chart shows the result of measurements at 5,000 cycles.

The voltage developed across the inductance was measured directly with a calibrated oscillograph and compared with the voltage developed across the ammeter shunt so that the inductive reactance could be calculated. This was then translated into inductance.

Resistance

The resistance measured at 1,000 cycles and multiplied by the amp-hour capacity rating of the battery is shown in Fig. 5. The smaller batteries show 0.14 ohms for the range of 10 to 100 amp-hours. In these batteries the metal parts are relatively large, their size being determined by mechanical need rather than by electrical resistance. Calculation shows that most of this battery resistance is caused by acid between the plates. Therefore, we have considered R_e as 0.14 ohm. This represents the acid path modified by the presence of separators, glass mats, and gas bubbles.

The acid resistance can be reduced, at the expense of other battery properties, by reducing the distance between plates, increasing the porosity of the separators, or reducing the number of layers of insulation.

For the batteries of 180 amp-hours or larger, the resistance per amp-hour is no longer constant, but increases gradually. It is believed that part of this reflects the larger size without a corresponding increase in the area of the grid ribs, and the larger number of plates, which require a longer strap to connect the plates to the post. Finally, the posts contribute significantly to the battery resistance in the larger battery sizes. These effects may amount to as much as 0.1 ohm for the largest battery sizes. This resistance can be reduced by the use of copper insert posts, and by a complete redesign of the battery if such a low resistance becomes essential.

Capacitance

When alternating current is passed through a battery, one component of the voltage lags the current 90 degrees, and can be balanced in the bridge circuit by a capacitor. The calculated battery capacitance is surprisingly large and amounts to 17,000 microfarads per amp-

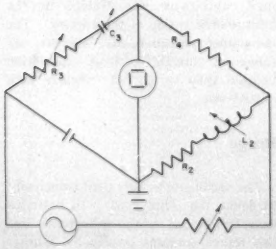


Fig. 3 (left). Wheatstone bridge

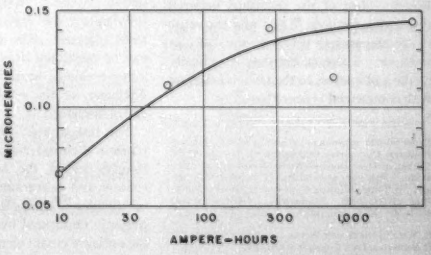


Fig. 4 (right). Inductance and battery size

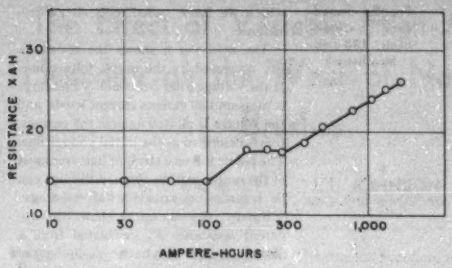


Fig. 3. Resistance and battery size

Table I. Measured Battery Capacitance

Amp-Hours	Capacitance, Farads	Farads per Amp-Hour
10	0.16	0.016
50	1.05	0.021
300	5.1	0.017
840	13.5	0.016

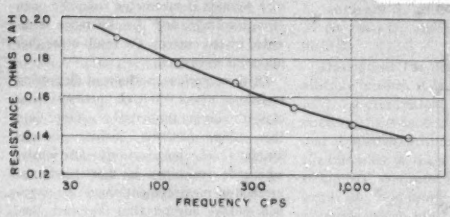


Fig. 6. Apparent resistance and frequency

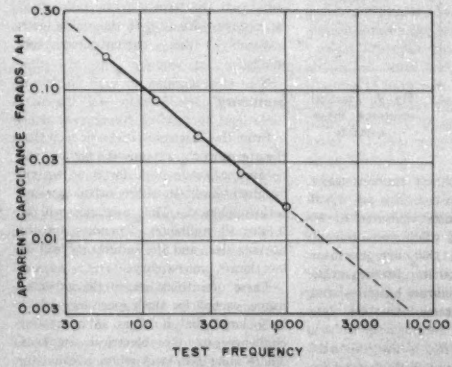


Fig. 7. Apparent capacitance effect of frequency

hour of capacity. A 1,000-amp-hour cell would, therefore, have a capacitance of 17 farads. This is several orders of magnitude larger than any other capacitor known, and offers a new circuit element for filtering or energy storage applications.

This capacitance is measured at 1,000 cycles. In view of the leakage resistance R_s the apparent capacitance becomes considerably greater at low frequencies. This will be discussed later.

Leakage Resistance

The battery capacitance describes only the initial slope of the voltage curve. The subsequent slope can be approximated by using the time constant of the circuit. The time required for 63% of the polarization to disappear amounts to about 3,000 microseconds, and this time constant corresponds to that of a 0.016-farad capacitor discharging through a resistance of approximately 0.20 ohm. Since this is on an amp-hour basis, the leakage resistance is therefore 0.20 ohm per amp-hour.

Where the discharge is at a high rate, and lasts for more than a second, this simplified picture no longer appears to be valid. The shape of the curve becomes more complex, and wide variations of the leakage resistance may be observed. This aspect of battery properties is not yet clear.

Where, instead of using pulse tests, the a-c bridge is used, the capacitance and leakage resistance form a frequency sensitive network which can dissipate energy. The apparent capacitance and resistance therefore, will vary with the frequency and since the equivalent circuit is not quantitatively correct, the variation with frequency must be measured.

Resistance and Frequency

The internal resistance of the battery has, up to this point, been the resistance measured at 1,000 cycles, and it has been found that this is identical with the resistance read from the pulse oscillogram of Fig. 2. If the test frequency is varied, part of the a-c energy is dissipated by the chemical reaction and shows up as an apparent increase in resistance as seen by the bridge circuit. This effect of frequency is shown in Fig. 6.

Capacitance and Frequency

The capacitance tabulated previously was the value measured at 1,000 cycles. When the test frequency is varied, the apparent capacitance varies, and the results can be plotted on a straight line up to 1,000 cycles as shown in Fig. 7. At higher frequencies a battery introduces too much inductance, but by using a specially designed test cell Lange has indicated that this line is straight to at least 20,000 cycles, provided data are plotted on logarithmic scales.

Relative Importance of Impedance Components

Up to this point the discussion has been primarily on an amp-hour basis in order to minimize the effect of battery size. The absolute values of the components of the battery impedance are shown in Fig. 8 for a 180-amp-hour cell. The experimental points fall on the solid line. Dotted extensions on these lines shown where we would expect the lines to fall if the areas were accessible to experimental work.

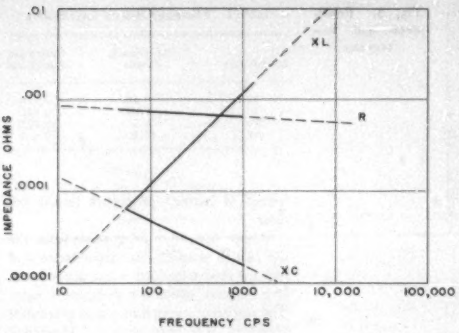


Fig. 8. Impedance components, 180-amp-hour battery

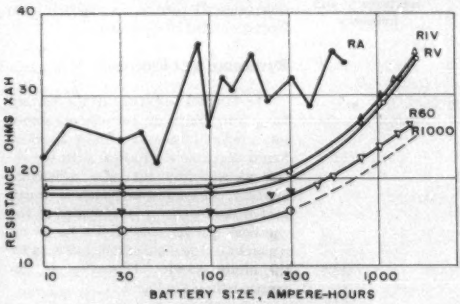


Fig. 9. Practical resistance measurements

It is clearly evident that the inductance of the battery controls its counter emf at high frequencies, and in view of the data in Fig. 4, the position of this line will not shift seriously with battery size. At low frequencies, both resistance and capacitive impedance control the counter emf. These vary inversely with battery size, and for an 18-amp-hour battery the values would be 10 times the quantities shown.

Practical Resistance Measurements

Various numbers associated with the electrical resistance are shown in Fig. 9. The bottom line shows the resistance measured by the a-c bridge at 1,000 cycles. The resistance measured at 60 cycles is shown next. This measurement avoids the need for a 1,000-cycle generator with an output of 100 amp for measuring the resistance of the larger batteries, and also makes inductance troubles less important.

The next curve, R_V, is the "virtual" resistance. If a battery is discharged for 5 seconds at a preselected current and the voltage is read immediately before, and 5 seconds after, the current interruption, the change in voltage divided by the current gives an apparent resistance which we have called the virtual resistance.

The next curve shows the resistance R_{IV} measured by the initial-voltage line. If the voltage after 5 seconds of discharge is measured at various current levels, and the voltage is plotted against the current, this is described as the initial voltage line. The points fall on a straight line over most of the range, and the slope of this line can be regarded as an electrical resistance.

Finally, the top curve shows the apparent resistance R_A calculated from a discharge pulse. If a battery is discharged for 5 seconds and the voltage at the end of 5 seconds is subtracted from the open circuit voltage, and the difference is divided by the current, the result is another apparent resistance.

Of these various methods of measuring resistance, the 1,000-cycle method comes closest to giving the actual resistance, and the other methods include various amounts of polarization. Obviously, whenever resistance is mentioned, the method of measurement should be specified and we suggest that the word, "resistance" be limited to the 1,000-cycle results, and the term "virtual resistance" be reserved for one of the pulse tests, preferably the current-interruption method.

Summary

From this discussion it can be seen that the apparent capacitance of a battery will be in the range of 1 farad to several hundred farads depending on the size and test frequency. The resistance will be 0.1 to 10 milliohms depending on the battery size, and the inductance will be less than 0.1 microhenry.

These quantities are in the expected range, except for the capacitance which is so large that it offers an interesting challenge to the electrical engineer. There must be places where a capacitor with these characteristics is needed.

Reference

1. Storage Batteries (book), G. W. Vinal. John Wiley & Sons, Inc., New York, N. Y., 1955

The Effect of Variable High-Altitude Humidity on the Wear of Nondusting Brushes

L. E. MOBERLY
NONMEMBER AIEE

J. L. JOHNSON
ASSOCIATE MEMBER AIEE

THE important effect which water vapor has on the performance of carbon brushes has been recognized for many years. Prior to 1935,[1] cases of excessively short brush life were found to be due to the low humidity of the cold air of midwestern United States winters. Early in World War II, when bombers began flying at altitudes above 20,000 feet, this low-humidity problem was more spectacularly demonstrated when it was found that the brushes on the dc generators were wearing out completely during flights as short as one hour. Tests in simulated high-altitude atmospheres showed, conclusively, that the extremely rapid wear, or "dusting," of carbon brushes was caused by the inability of the brushes to maintain lubricating films on commutators in the extremely low atmospheric humidities which were found to exist at high altitudes.

It became necessary to develop a new type of brush for high-altitude service. These so-called "altitude brushes" contain special constituents which substitute for atmospheric water, providing low-friction running and acceptable brush life even at extremely low humidities.

The earliest special-altitude brushes which were successfully applied were the metallic halide containing brushes of Dr. H. M. Elsey. Theories providing possible explanations of the role which water, or its substitutes, plays in maintaining lubricating films between carbon brushes and commutators are discussed by Elsey,[2] Savage,[3] Sims,[4] and Campbell.[5] This paper, however, is concerned not so much with the mechanism by which altitude treatments function, but rather with the actual performance of presently available high-altitude brushes under conditions which may be encountered in modern flights.

Throughout the development of high-altitude brushes, it has been necessary to rely to a great extent on evaluation data obtained from laboratory bell jar testers and simulated-altitude chambers. Since the humidity of the atmosphere has been recognized as a most important factor in brush life, this is one of the testing parameters which has been maintained at levels consistent with conditions anticipated at altitudes of actual flight. Information concerning actual humidities at various altitudes has been difficult to obtain but, until fairly recently at least, it seems to have been generally accepted that the humidity represented by a dew point of −50 C (degrees centigrade), or lower, has been acceptable for evaluating altitude brushes.

Fig. 1 shows the relationship between dew point and absolute humidity. The absolute humidity of air whose dew point is −50 C is shown to be 0.02 grain of water per cubic foot. This is a very low water content, particularly when compared with the humidity of 2–4 grains per cubic foot which is normally supplied to brushes operating at the earth's surface. It has frequently been assumed that 0.02 grain per cubic foot represents such a low level of water content that drying to still a lower humidity would produce a negligible change in the performance of carbon brushes.

Along with the development of the jet airplane, with its capabilities of flying higher and faster, the performance requirement for altitude brushes became even more severe. Where the aircraft and its equipment formerly were operated for appreciable periods at sea level, they could now be taken almost immediately to high altitudes. This meant that the brushes must be capable of performing satisfactorily under high-altitude conditions with negligible prefilming of commutators or slip rings. Brushes which meet this requirement are currently available and have become known as "immediate-filming" altitude brushes.

A number of immediate-filming brush grades have been successfully tested during the past few years in the Westinghouse laboratory bell jar tester at an established testing condition simulating air pressures at 60,000 feet altitude and at a dew point of −60 C. Some of these grades have also been operated satisfactorily on electric equipment in altitude chambers in which dew points are reported to be −50 C, or lower. Following the successful evaluation tests, as is the usual practice, these brushes have been applied on aircraft for actual high-altitude service and, in general, have shown excellent performance.

Recent data, however, indicate that altitude brushes might experience occasional erratic and unexplained accelerated brush wear on actual flight tests. Although the wear in these cases was much too low to be considered dusting, it might be several times that expected from testing experience when taking into account the altitude service which has been endured.

A study of available meterological sounding data[6,7] indicates that the absolute humidity of air between 20,000 feet and 65,000 feet probably is quite variable depending to some extent on geographical location and to seasonal fluctuations. In fact, there is good evidence that dew points of the air at 60,000 feet, for example, may vary as much as from −50 C to −80 C.

It occurred to us that although the difference in the total water content of air with −50 C dew point and −80 C dew point is extremely small, altitude brushes

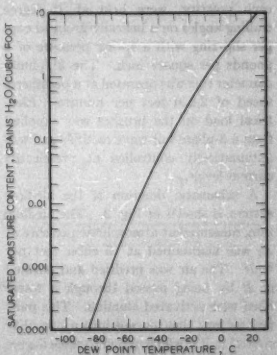


Fig. 1. Grains of moisture per cubic foot (saturated gas) versus dew-point temperature

Paper 59-565, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted August 1, 1958; made available for printing February 24, 1959.

L. E. MOBERLY and J. L. JOHNSON are with the Westinghouse Electric Corporation, Pittsburgh, Pa.

The authors wish to acknowledge the advice and counsel of Dr. H. M. Elsey, chemical consultant, who originally foresaw the possibility of the effects pointed out in this paper.

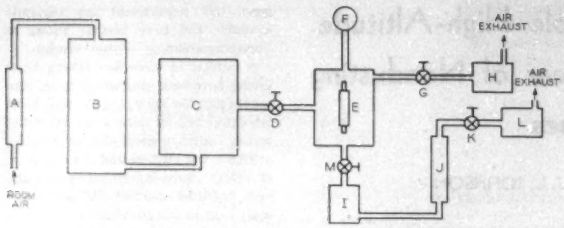


Fig. 2. Schematic of components comprising air-flow system

A—Flow meter
B—Air predryer and purifier
C—Air-humidity control
D—Constriction
E—Test chamber, slip ring, brushes
F—Vacuum gage
G—Valve
H—Vacuum pump
I—Dew-point cell
J—Flow meter
K—Valve
L—Vacuum pump
M—Valve

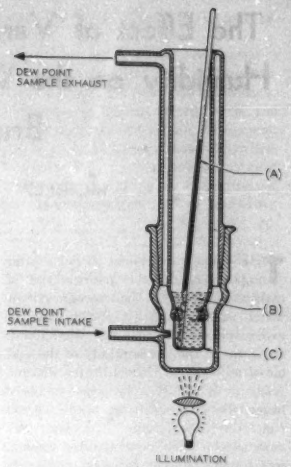


Fig. 3. Cross section of dew-point cell

A—Thermometer
B—Cooling medium
D—Mirror polish cup

still might be sensitive enough to be affected appreciably by this difference. If this were true, the erratic brush wear which has been experienced on actual altitude flights could be explained by variable atmospheric humidity.

Apparatus and Techniques

A laboratory bell jar tester, employing all the usual precautions for preventing contamination from organic materials and other sources, was utilized for this investigation. The tester has been generally described in an earlier paper[8] and only slight modifications in the air-flow system were necessary for the present high-altitude tests. The brushes and brush holders were the same design as are currently applied on the exciters of some aircraft alternators. Two brushes, one of each polarity, were held at 15-degree trailing angles on a helically grooved copper slip ring with a spring pressure of 6 pounds per square inch. The $3^1/_4$-inch-diameter ring was operated at a peripheral speed of 2,930 feet per minute. Electrical load on the brushes was supplied from a 3-phase full-wave rectifier and was automatically controlled at preselected current levels.

A schematic diagram of the air-flow system is shown in Fig. 2. The air-flow rate, measured at atmospheric pressure at *A*, was maintained at 10 cubic feet per hour. The air was predried and purified at *B* by being passed through a tower filled with activated alumina. This purified air was then passed into a humidity control unit *C* which varied to some degree depending upon the humidity level desired for a particular test. A constriction at *D* expanded the air to a pressure equivalent to 60,000 feet altitude inside the test chamber *E*, measured by gage *F*, and maintained by vacuum pump *H* through a control valve *G*. A dew-point measuring apparatus *I* was connected to the test chamber. A measured flow of air, using flow-meter *J*, could be drawn from the test chamber through the dew-point apparatus using vacuum pump *L*. This air constituted a small fraction, 5%, of the total air flowing through the system. Thus, a direct measurement of the humidity existing in the test chamber could be made at any time during the test.

Humidities were measured by the dew-point cup method which has been found by experience to give very reliable measurements, particularly in the range studied.[6,9] The actual apparatus used, shown in Fig. 3, is one of the many modifications of cooled cup devices which can be used for measuring dew points. An important feature of this particular apparatus is that it can be used at the low air pressures which exist at high altitudes. The techniques utilized in obtaining the dew point of the air were those commonly practiced. A dry-ice–acetone mixture was used as the coolant and the temperature was measured with a calibrated, low-temperature thermometer.

The observation of the dew point, or frost point, was enhanced through the use of special illumination on the mirror. The amount of moisture present in the air at

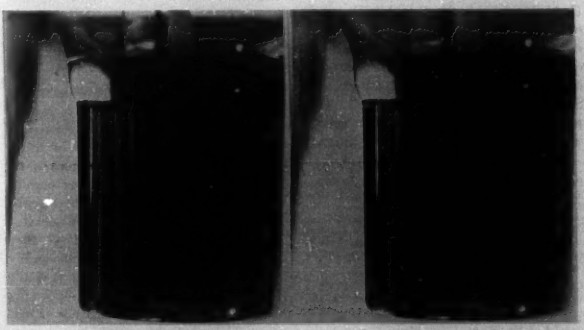


Fig. 4. Appearance of dew-point mirror before (left) and after (right) frost formation

Table I. Comparison Between Calculated and Measured Humidities with Westinghouse Dew-Point Cell

Condition of Air							
Before Pressure Reduction			After Pressure Reduction				
	Measured Humidity			Calculated Humidity		Measured Humidity	
Pressure, Millimeters Mercury	Dew Point, C	Grains per Cubic Foot[3]	Pressure, Millimeters Mercury	Dew Point, C	Grains per Cubic Foot[3]	Dew Point, C	Grains per Cubic Foot[3]
719	+17	6.30	54	−18	0.474	−19	0.430
719	0	2.15	54	−29	0.161	−31	0.135
719	−28	0.18	55	−52	0.0138	−51	0.0155
719	−44	0.034	55	−64	0.0026	−64	0.0026
719	−47	0.024	19.8	−74	0.00065	−74	0.00065

dew points of −60 C, or below, is quite small and thus very light deposits of frost are obtained. It has been found that at these dew points the moisture condenses on the mirror as individual minute ice crystals which can be observed, readily, only through their reflection of light from special light source. Fig. 4 compares the appearance of a dew-point mirror before frost forms and at a dew point of −74 C.

The data in Table I were taken to show the reliability of the dew-point readings taken with our apparatus. In this case, air was passed through a completely sealed glass system and its humidity measured, first at room pressure and again after expansion to a low pressure. By measuring the pressure of the air at both conditions, it was possible to calculate what the humidity should be at the expanded pressure from the easily measured and controlled value obtained at room pressure. The table shows quite good agreement between measured and calculated dew points indicating very good reliability for our measuring apparatus and techniques in the range of humidities investigated.

For this investigation on brushes, it was desirable to test at several humidity levels while maintaining the air pressure, or "altitude," at a fixed 60,000 feet. This was accomplished by making variations in the air humidity control unit referred to previously. For example, to provide a humidity of −20 C at 60,000 feet altitude, air was saturated with water at +17 C at room pressure in the air-humidity control unit. The ultimate expansion of this air to the 60,000 feet altitude pressure resulted in an atmosphere whose dew point was −20 C. Similarly, air saturated at 0 C at room pressure, when expanded, gave dew points in the order of −30 C to −35 C. For humidities in the range of −55 C dew point, the air was saturated at −30 C before expansion. For the lowest humidities tested, the air was passed through a dry-ice-cooled tower prior to expansion to 60,000 feet altitude. Resulting dew points in the range of −65 C to −80 C were measured.

It should be pointed out that the calculated humidity of air which is saturated at dry-ice temperature and then expanded to a 60,000-foot altitude is considerably lower than the −65 C to −80 C which was obtained. The difference can be easily explained by the occurrence of extremely small and variable leakage of room air into the low-pressure system. This leakage is one of the problems universally associated with simulated altitude testing and is the reason why humidity must be measured in the test chamber itself. We have had the opportunity of measuring humidities in several altitude chambers in which it was customary to calculate "chamber humidity" from the values measured on the ventilating air prior to expansion. In every case, considerably higher humidity values were found than had been assumed. Calculated air humidities can be reliable

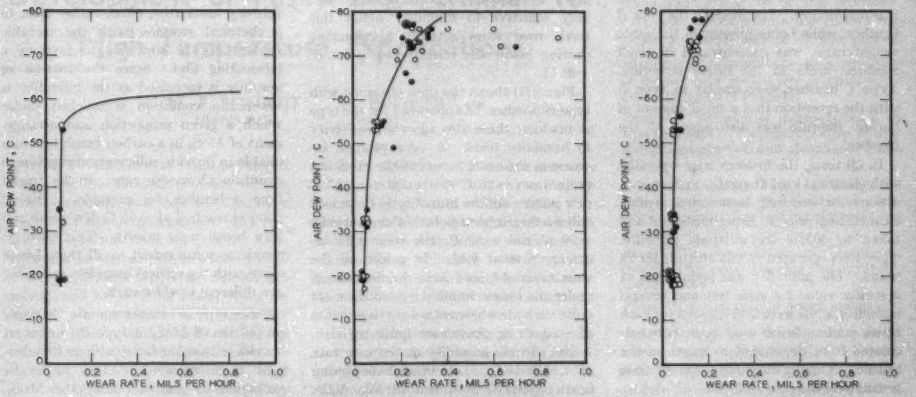


Fig. 5. Wear rate of brushes as a function of air humidity

A—Type A brushes
B—Type B brushes
C—Type C brushes
Operating conditions: 60,000 feet altitude, 6 pounds per square inch, 2,930 feet per minute
● 107 amperes per square inch
○ 60 amperes per square inch

only if the system is absolutely tight since even a small leakage of moist room air into the extremely dry test chamber will produce a major change in the ultimate humidity level.

For example, a normal value for room air humidity is 50% relative humidity at 25 C. This corresponds to an absolute humidity of 5 grains per cubic foot. The absolute humidity of air whose dew point is −80 C is 0.00026 grain per cubic foot. Thus, when these two are mixed, as occurs in a leaking "altitude" system, the humidity ratio involved is 20,000 to 1, i.e., one volume of room air leaking into 20,000 volumes of dry chamber air would double its water content.

Experimental Results

The brushes tested were commercially available grades of the immediate-filming type. The principal filming agent in all the brushes was molybdenum disulfide (MoS_2), which has become established as the most generally applied material for obtaining high-altitude protection without prefilming. The brushes studied were conveniently separated into three types based essentially on the manner in which the MoS_2 was applied. All three types contained the same weight ratio of MoS_2 to carbon in the working section of the brush. In brush type *A*, the MoS_2 existed as extremely fine particles which were uniformly distributed throughout the brush structure. The MoS_2 in the type *B* brushes, while being present in the same weight ratio, was concentrated in small localized areas in the brush structure. Type *C* brushes were similar to type *B* with the exception that a small amount of barium fluoride was also added to the brushes.

In all tests, the brushes were operated with electrical load at earth's surface conditions for one-half hour on a freshly cleaned ring prior to being measured and taken to 60,000 feet altitude pressure. They were operated at this altitude for 24 hours. The humidity was maintained at a steady value for each test and several humidity levels were investigated for each brush grade. Brush wear rates were calculated from the change in length of the brushes occurring during the high-altitude testing period.

Since the electrical load on brushes may vary considerably in actual flight service, it was interesting to determine what effect load conditions might show on brush wear. Two loads, 60 amperes per square inch and 107 amperes per square inch were compared as representing typical load variations. It can be seen from Fig. 5 that no detectable difference in wear occurs within this load range.

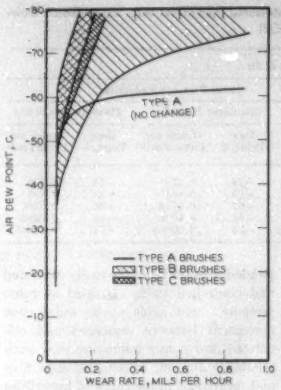


Fig. 6. Brush-wear rate range as a function of air humidity

60,000 feet altitude, 6 pounds per square inch, 60–170 amperes per square inch, 2,930 feet per minute

Fig. 5(A) is a curve showing the relationship between brush wear rate and humidity for brushes of type *A*. These brushes appear to operate uniformly well at humidities represented by dew points down to −50 C. The brushes become very sensitive to humidity below this level, wear rates actually approaching dusting below the critical dew point of −55 C.

Fig. 5(B) shows the data obtained with type *B* brushes. As observed for the type *A* brushes, these also show a sensitivity to humidity level. A very gradual increase in wear rate occurs as the humidity drops from −20 C dew point to −55 C dew point. As the humidity is decreased still further to dew-point values between −70 C and −80 C, the wear rate increases several fold. In addition, the wear rates obtained from duplicate tests under the lowest humidity conditions are quite variable whereas those at dew points of −55 C, or above, are quite uniform.

The effect of humidity on the wear rate of a third type of immediate-filming brush, type *C*, is shown in Fig. 5(C). Here again, the increasing wear rate occurs at humidities lower than −55 C dew point. In this case, however, the wear rates are still reasonably low even under the lowest humidity conditions which were studied.

Fig. 6 shows the combined test results from all three brush types. Here the range of wear rates obtained for a large number of tests on each brush type is shown as a shaded area. A comparison between the different types shows that, in respect to wear rate, an improvement in brush life is gained in going from type *A* through type *B* to type *C*. Wear rates for type *C* brushes are lower and much more consistent than was found for the other two types. No area is shown for the type *A* brushes since very little range was observed in their wear performance. Down to −50 C dew point, the values were quite uniform from test to test, while at lower dew points the wear rate was extremely rapid. Thus, the humidity level near −55 C dew point appears to be quite critical for brushes of this type and they are expected to be unsatisfactory for extended periods at humidities at or below this level. Type *B* brushes fall somewhere between the others, becoming somewhat erratic in wear at humidities lower than −70 C dew point.

Since the geometry of the added MoS_2 is the major difference between type *A* and type *B* brushes, it is apparent that this is important in providing the differences in wear characteristics noted between the two types. No positive conclusion has been developed to explain why this is so but it is interesting to conjecture.

MoS_2 is a good dry lubricant in vacuum[10] in the absence of moisture. It is presumed that the low-humidity protection of carbon brushes by MoS_2 is due to this dry lubricating effect rather than to a chemical reaction with the metallic commutator or slip ring to produce a lubricating film. Since the friction of graphite is increased as the humidity is lowered a condition is reached under which a given proportion and arrangement of MoS_2 in a carbon brush becomes unable to provide sufficient lubrication to maintain a low wear rate. In the case of type *A* brushes, for example, a critical point is reached at −55 C dew point and high brush wear results. This effect is shown to some extent in all three brush types with the critical humidity occurring at a different level for each.

Since type *A* brushes contain the same proportion of MoS_2 as type *B* brushes, it is evident that the latter utilizes its lubricant more efficiently. This might be explained by the fact that the MoS_2, which exists in type *A* brushes in a particle size of the same order as brush wear particles, is uniformly distributed in the carbon network in the brush. This effectively dilutes the MoS_2 over the full contact area of the brush. As brush wear occurs, a relatively high percentage of these MoS_2 particles become thrown off

with the wear dust and thus cannot be transferred to the brush track to help maintain a low-friction film.

In contrast, in the type *B* brushes, the large MoS_2 particles are held firmly in the carbon structure. As each particle becomes exposed at the brush face through normal brush wear, it is rubbed directly onto the ring surface and thus a heavily filmed area is developed on the brush track. With the random distribution of MoS_2 particles, practically all areas of the brush track receive the benefit of the filming action. Since the mechanical load is primarily supported by these heavily filmed areas, the net result is lower wear and better altitude protection for the type *B* brushes.

The improvement exhibited by type *C* brushes can be explained by the extra "filming" obtained from the barium fluoride whose properties as a high-altitude brush treatment are well known. The effects of MoS_2 and BaF_2 appear to be additive.

Although this study has been extremely useful in the evaluation of brushes for Westinghouse aircraft equipment, it is recognized that there are many characteristics of brushes, in addition to wear rate, which must be considered when making selections for individual machines. Therefore, it should be emphasized that brushes of types *A*, *B*, or *C* are not being ranked here for any specific application on the basis of the foregoing data. The chief purpose of this particular study has been to show that all altitude brushes are affected by variable atmospheric humidity and that the level at which this becomes critical varies from one type of brush to another. Even the best brushes show wear rates at −80 C dew point which are four times those at −55 C dew point.

Summary

It is well known that untreated carbon brushes show dusting wear at high altitudes due to the existence of extremely low air humidity. Even though modern, specially prepared, altitude brushes may not experience this disastrous dusting wear, evidence is presented in this paper which indicates that their wear rates are a function of the absolute humidity of the high-altitude air. Previously unexplained cases of erratic brush wear on actual flights can now be explained by the range of fluctuations which may occur in the humidity of the atmosphere at present flight altitudes.

References

1. The Effect of Humidity on Brush Operation, J. V. Dobson. *Electrical Journal*, East Pittsburgh, Pa., vol. 33, Dec. 1935, pp. 527–28.

2. Treatment of High-Altitude Brushes by Application of Metallic Halides, Howard M. Elsey. *AIEE Transactions* (*Electrical Engineering*), vol. 64, Aug. 1945, pp. 576–79.

3. Graphite Lubrication, R. H. Savage. *Journal of Applied Physics*, New York, N. Y., vol. 19, Jan. 1948, pp. 1–10.

4. The Wear of Carbon Brushes at High Altitudes, R. F. Sims. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. I, July 1953, pp. 183–88.

5. Wear of Carbon Brushes in Dry Atmosphere, W. E. Campbell, R. Kozak. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 70, July 1948, pp. 491–98.

6. Meteorology of the Lower Atmosphere, G. M. B. Dobson, A. W. Brewer, B. M. Cwilong. *Proceedings*, Royal Society of London, London, England, series A, vol. 185, Jan.–Apr. 1946, pp. 144–75.

7. Some Measurements of the Distribution of Water Vapor in the Stratosphere, Earl W. Barrett, Lee R. Herndon, Jr., Howard J. Carter. *Tellus*, Stockholm, Sweden, vol. 2, 1950, pp. 302–11.

8. Air Humidity and Brush Contact Drop, Howard M. Elsey, Lawrence E. Moberly, John L. Johnson. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 73, Dec. 1954, pp. 1383–89.

9. The Measurement of Water Vapor in Air and Other Gases, Elmer R. Queer, E. R. McLaughlin. *Bulletin no. 224*, Pittsburgh Lectrodryer Corporation, Pittsburgh, Pa.

10. Investigation of MoS_2 Lubrication in Vacuum, Virgil R. Johnson, George W. Vaughn. *Journal of Applied Physics*, vol. 27, no. 10, Oct. 1956, pp. 1173–79.

Development of Fuses and Terminals for High-Temperature Applications

W. F. BONWITT
MEMBER AIEE

H. BUTTNER
MEMBER AIEE

COMPONENTS TO BE developed under the Hotelec program include fuses, terminals, and splices. The most significant new requirement of this program as it affects these components is the extension of the temperature level to 850 F (degrees Fahrenheit) for terminals and splices, and to 600 F for fuses; these are ambient temperatures with conductor temperatures reaching 1,000 F. Other significant requirements are for the devices to withstand vibration to 20 *g* (gravitational acceleration) at a frequency of 80 to 2,500 cps (cycles per second); acoustic noise of 150 decibels, as well as shock of 50 *g* for 11 milliseconds. The altitude of 80,000 feet demands resistance to ozone, and requires consideration of the effects of corona. Similarly, the significant increase of temperature makes it necessary to limit thermal voltages at junctions to 0.1 millivolt.

It was felt that the increased temperature was very likely to be the most rigorous new requirement. Therefore, the materials used in present designs were analyzed. In the case of terminals and splices, which are preferred in the preinsulated, compression type, it soon became obvious that new materials or techniques would have to be developed. In the case of fuses, the new temperature requirement made it necessary to replace plastics with ceramics. New methods and materials to join them to the metal components of the fuse also became necessary.

Hotelec developments are carried out under a general scheme of five phases:

Phase I: analysis of requirements, the state of the art, and preliminary material selection. Planning of Phase II.

Phase II: selection of materials based on screening tests. The determination of suitable design concepts and planning of Phase III.

Phase III: construction of breadboard models and tests to prove out their feasibility. Specification for prototypes and planning of Phase IV.

Phase IV: construction of prototypes. Qualification tests following the applicable specifications.

Phase V: preparation of design manuals.

At this time both developments are at various stages of Phase III. The problems encountered and the progress made to date are described separately for fuses, and for terminals and splices.

Terminals and Splices

The most direct approach to the achievement of high-temperature terminals and splices would be the substitution of high-temperature materials for those

Paper 59-771, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 13, 1959; made available for printing April 8, 1959.

W. F. Bonwitt and H. Buttner are with the Burndy Corporation, Norwalk, Conn.

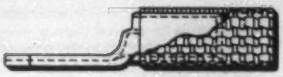


Fig. 1. Preinsulated terminal

used in existing designs which are generally limited to temperatures below 400 F. Present terminals consist of a conducting body made of copper or copper alloy suitably plated and an insulating sleeve made of a thermoplastic material such as vinyl, nylon, or Kel-F. While there is little doubt that materials and finishes suitable for the conducting body exist and require only evaluation and selection, the same can hardly be said for the insulating sleeve. Materials which will withstand the plastic deformation under load incident to crimping and which will provide wet dielectric strength throughout the temperature range −65 F to 1,000 F do not exist and are not readily discernible on the technological horizon. Thus, the problem of preinsulating constitutes the main challenge to the designer of high-temperature terminals and splices.

Over 30 means of attaining preinsulated terminals have been considered. Most of them have been discarded in the conceptual stage when measured against the practical yardsticks of size, weight, reliability, and the requirement for field maintenance installation tools. Following are some which have been considered extensively:

Fig. 1 shows a conventional terminal body with a fiberglass insulating sleeve. Crimping imposes a severe stress on the sleeve. Single, double, and triple layer sleeves, each layer 0.015 inch thick, were employed and the resultant assemblies subjected to a dry dielectric test. Marginal results were obtained with single and double sleeves but triple layer sleeves provided dielectric values of 2,400 volts, well in excess of the 1,200 volts considered a minimum test value. Of course, it is highly desirable to provide wet dielectric strength in terminals and splices since they will be subject to humidity and may be in contact with each other or grounded material in any practical application. This leads to the search for a material to impregnate the fiberglass and make it water repellent. Such material must have enough tenacity or affinity for the fiberglass to remain in sufficient quantity under the crimp to carry out its assigned function. It must not carbonize or otherwise act as a conductor at any temperature up to 1,000 F and must retain its water-repellent quality up to this temperature.

A second approach to the insulating sleeve problem is the use of a heavily anodized aluminum sleeve. Anodized aluminum wires have reputedly been used successfully in high-temperature transformer and motor windings. The voltage between adjacent turns or layers is of course much lower than that which is required for terminals and splices. There is considerable doubt as to the ability of such coatings to withstand the crimping forces and the dielectric strength required. Anodized coatings are porous and wet dielectric strength is questionable of attainment. This is therefore not considered a very promising approach.

Flexible ceramics have been suggested. These materials are flexible to a limited degree and then only in the "green" state in which they are also highly water absorbent. Thus their use would require curing after crimping, hardly feasible for field maintenance.

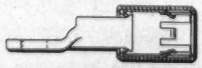
While the prospects for a preinsulated terminal having wet dielectric strength are not too promising, there are several alternate courses to be pursued.

Fig. 2 shows what might be termed an almost preinsulated terminal. Crimping is done directly on the terminal barrel itself. The captive insulating sleeve is then pushed forward until it locks in place. This approaches the convenience of a preinsulated terminal. The terminal size can be kept within reasonable limits if a ceramic-clad metal sleeve is used and if crimping is such as to keep the terminal barrel within its original outline. Such crimping is attained without undue complexity in tools. This design is suitable for splices as welll as terminals. A limitation is that insulation grips are not readily incorporated without adding extra length. The sleeve retainer would have to be approximately 1/8 inch away from the end of the grip-closing die during installation.

Sometimes a difficult problem can be met by eliminating it. Fig. 3 shows a termination which is so compact as to eliminate the need for insulation. Obviously the tongue thickness becomes excessive for large size conductors. However it seems quite practical as to size and economy for wire sizes No. 16 and smaller. Such wires constitute at least 90% of any aircraft or missile electrical system. Insulation grips also seem feasible in these small sizes. This concept is obviously not suited to splices. A possible compromise solution of the over-all problem is the use of Fig. 3 for small size terminals and the use of Fig. 2 for larger terminals and all sizes of splices since insulation grips are not essential for satisfactory performance of splices and large size terminals. This solution would entail the use of duplicate crimping tools for small size terminals and splices. Therefore, the search for a solution which is applicable to both terminals and splices in a given size, or which offers a more acceptable compromise, continues.

The determination of suitable materials and finishes for the conducting body of a terminal or splice is not nearly so formidable. In Phase I the known characteristics of various metals were reviewed. Nickel, silver, and copper were selected for further investigation.

The type of nickel selected for further study is known as *A* nickel, electronic grade. This nickel has about 27% of the conductivity of copper and has a coefficient of thermal expansion very close to that for steel. It is magnetic at temperatures up to 680 F, having a maximum permeability, μ at 20 C (degrees centigrade) of 600 as compared with iron which has a value of 5,000. From its modulus of elasticity and thermal expansion it was



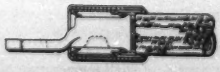
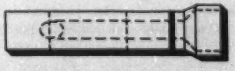


Fig. 2. Post-insulated terminal



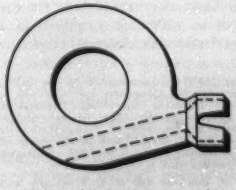


Fig. 3. Washer-type terminal

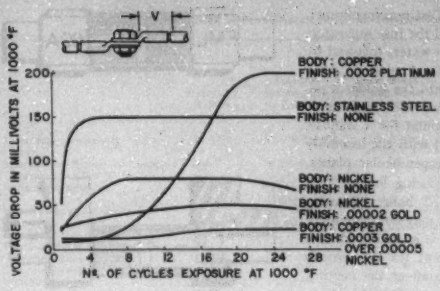


Fig. 4. Effect of high - temperature cycling on various terminal materials and finishes. Each cycle consisted of 7 hours at 1,000 F with 46 amp, and 17 hours of room temperature and no load. Conductor was MT-6 as specified in reference 1

predicted that nickel will not creep under clamping pressures which would normally be encountered. This was borne out in Phase II screening tests in that the initial tightening torques for stainless steel bolts used to clamp together terminals made of nickel, did not change after repeated exposures to temperatures of 1,000 F. Its lower conductivity relative to copper was reflected in these tests by a higher voltage drop. This can be partially offset by plating with gold which has a lower contact resistance and, being softer, provides more points of contact and therefore greater effective contact area. The magnetic properties of nickel do not seem to result in undue heating or other undesirable effects in the applications intended. Nickel-clad copper conductors as described in reference 1 have been used for some 10 years in aircraft electrical systems without reported difficulty.

It was suggested that the high-temperature properties of silver would be improved when alloyed with small percentages of magnesium and nickel. It was found however that the conductivity of this material changes markedly after even a few cycles of exposure to high temperatures. The voltage drop at rated current increased tenfold after only 6 cycles. Thus, this material was dismissed from further consideration.

As regards copper, only the type known as OFHC (Oxygen Free High Conductivity) was considered. Its surface must be protected because of its high oxidation rate at elevated temperatures. Its modulus of elasticity and coefficient of thermal expansion are such as to introduce a question of the extent to which it will creep under elevated temperatures. As measured by the starting torque required to retighten a stainless steel bolt and nut clamping two such terminals together after each exposure to elevated temperature, the creep is sufficient to cause a drop in torque of approximately 50%. At the same time, retightening such bolted assemblies did not materially alter the voltage drop across the bolted connection. Thus, the type of difficulty encountered in the use of aluminum at substantially lower temperatures is not anticipated here.

The finishes considered were nickel, silver, gold, platinum, and rhodium. Different thicknesses were studied and in some cases nickel was used as an undercoat. The possibility of a stainless steel cladding on copper was also studied.

It had been reported that nickel could develop extremely high contact resistance after exposure to high temperature. This was observed in the case of a thin (0.0001/0.0002-inch) nickel plating on copper. After 4 and 5 cycles of high-temperature exposure it required 18 and 25 volts respectively to pass current across the interface of two such surfaces. This was not observed with nickel plating of approximately 3 times this thickness. In fact, the "starting" currents for all other materials and finishes tested were below 4 microvolts, which was the limit of sensitivity of the instruments used.

Gold appears to be the most satisfactory finish for this application. Fig. 4 shows the effect on voltage drop of repeated exposure to high temperature. Gold alone over copper permits oxidation of the copper, particularly where the plating has been thinned under the crimp. A nickel under-plate provides a degree of protection against this oxidation.

The requirement that the thermal voltage between the terminal and the conductor due to the thermocouple effect shall not exceed 0.1 mv (millivolt) through a conductor temperature range of −65 F to 1,000 F, seems to impose a severe restriction on the choice of material. The thermal voltages generated by several metal combinations at 1,000 F referred to a similar cold junction at 32 F are as follows:

Junction Metals	Thermocouple Voltage, Mv
Copper/Nickel	13.5*
Copper/No. 304 Stainless Steel	3.13†
Nickel/No. 304 Stainless Steel	−9.86†

* From American Institute of Physics Handbook, McGraw-Hill Book Company Inc., New York, N. Y., 1957.
† Data furnished by Lewis Engineering Co., Naugatuck, Conn.

While it is likely that a temperature difference of approximately 1,000 F may exist between the ends of the conductor, the temperature gradient across the terminal itself is very small. Assuming a nickel terminal on copper wire, an unlikely temperature gradient of 35 F across the terminal would be required to produce a thermocouple potential of 0.1 mv. The conductor will very likely consist of a copper core clad with nickel. While this combination would undoubtedly complicate the computation of junction potentials it cannot help but further improve a presently tolerable condition.

Consideration of the generated thermal voltages listed above leads to the suspicion that a problem may exist where the aircraft structure is used for grounding since this structure will be at various temperatures at any given moment. Such considerations are however outside the scope of the development of high-temperature terminals and splices.

High-Temperature Fuse

The fuse is to be self-indicating, with one size (configuration) accommodating several current ratings from 1 to 60 amperes. The fuse should also be as small and light as possible, and shall be distinct from present low-temperature fuses so that any chance of accidental mix-up is eliminated. As specified now and shown in the time-current relation Fig. 5, the primary application will be for system protection. However, wire protection may become a requirement later on, making a change of the time-current characteristic curve necessary. The fuses

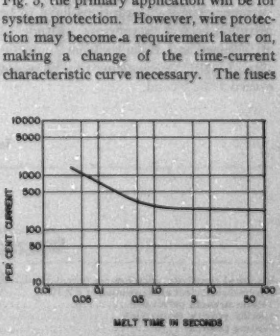


Fig. 5. Time-current characteristic

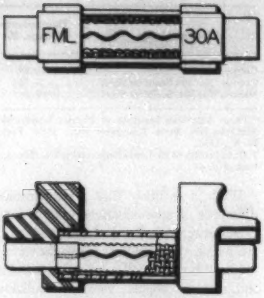


Fig. 6. Visually indicating fuse

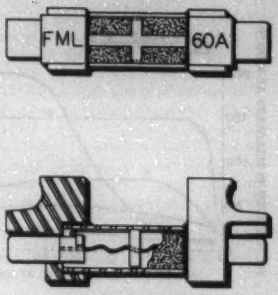


Fig. 7. Visually indicating fuse with separate indicator

are to be used in 3-phase 115/208-volt 400-cps electrical distribution systems so that the various electrical properties of presently available low-temperature fuses will generally remain unchanged.

The test requirements for the fuses are listed in reference 2. Some of the significant tests follow.

A Life Test requires the fuses to carry rated current at an ambient temperature of 100 F for 300 hours at sea level. This is followed by a 100-hour test, divided into 10 cycles of 8 hours at 100 F and 50,000-foot pressure altitude, and 2 hours at 600 F at 80,000-foot pressure altitude. Finally the fuses are exposed for 100 hours at 100 F at sea level. No change of characteristic may result from this test.

A Temperature Shock Test requires the fuses to pass the temperature range from −65 F to 600 F at a rate of 150 F per minute, carrying rated current.

The fuses have to interrupt a maximum current of 4,000 amp (amperes) line-to-ground at 115 volts, 400 cps; a maximum current of 2,500 amp line-to-line at 200 volts, and 4,000 amp at 28 volts direct current. These interruptions have to be accomplished at a pressure altitude of 80,000 feet and at an ambient temperature of 600 F. The arcing time may not exceed 0.002 second.

A Wet Dielectric Test requires immersion of the holder and the fuse separately for 10 seconds in tap water, followed by unagitated drainage for 1 minute. After the fuse is inserted into the holder a potential of 1,200 volts has to be maintained between fuse and ground for 1 minute. The test is performed with the assembly in three mutually perpendicular planes.

When tested after being loaded with rated current for 1/2 hour, the time-current characteristic may not deviate by more than 10%, nor is an overload of 5 times the rated current for 0.1 second, repeated 10 times, to affect the melting characteristic of the fuse. It is of interest that during all tests that may affect the fuse by the high ambient temperature, the fuse and holder must be mounted on a titanium panel 0.05-inch thick.

Two types of self-indicating fuses for temperatures to 250 F are available today: mechanically indicating fuses, and fuses where the interrupted fuse wire itself, or a separate indicator wire in parallel with the fuse wire, acts as a visual indicator. Since visually indicating fuses are much simpler, having no moving parts, it seemed desirable to select this type for high-temperature application.

Visually indicating fuses consist of a tubular translucent housing made usually from a Pyrex glass as shown in Fig. 6. Cemented to the ends of this housing are two molded heat-resistant plastic end caps. These caps act as supports for the blade contacts that terminate the fuse wire. The blades connect to corresponding contact clips in the separate fuse holder or base. The caps also provide convenient finger grips to insert and remove the fuse from its base without the use of tools. A spent fuse is recognized by the break in the fuse wire visible through the glass housing. In larger fuse ratings the considerable arc energy released under certain operating conditions makes it necessary to fill the housing with an arc-quenching filler as shown in Fig. 7. This makes the fuse element invisible, and in such ratings an indicating element in parallel with the fuse wire is placed adjacent to the inside wall of the glass tubing making it readily visible from the outside.

The fuse holder is usually a molded block accommodating fuses of the same physical size, but not necessarily of identical ratings. Open contact clips engage the blade contacts of the fuse, and terminate in contact pads with conventional hardware for attaching wire terminals.

The heart of the fuse is a properly selected fuse wire assembly. Two fuse wire metals are used today: silver for the larger ratings, and nickel for ratings below 5 amp. The reason for using nickel in the small current ratings is one of mechanical strength; silver wire of the required cross-section would be so slender that it would be difficult to handle, quite sensitive to damage, and not readily visible. Nickel has a higher resistivity and a time-current characteristic similar to that of silver. Its cross section and its strength are therefore greater than that of silver for the same fuse rating. Whatever the metal of the fuse wire, it is silver soldered to silver plated copper contact blades.

Considering first the fuse wire, the melting temperature of silver is 1,760 F, and that of nickel 2,642 F. Both are considered sufficiently above the 1,000 F conductor temperature to make them

Table I. Properties of Metals for High-Temperature Fuses

	Gold 99.99+[3]	Copper OFHC[4]	Silver 99.97% Pure[3]	Nickel Inco Grade A[3]	Easy-Flo 45[5]	Rhodium Pure[3]	Inconel X #1 Temper[7]	Stainless Steel Alloy 416 Heat-Treated[8]	Beryllium Copper Berylco Alloy 10 1/2 Hard Heat-Treated[3]
Electrical resistivity microhm-cm (circular mils)	2.35	1.71	1.59	10.0	6.25	4.5	124.6	57.0	3.8–2.9
Thermal coefficient of expansion per F×10⁻⁶	9.9	9.4–9.8	10.9	8.3	10.0	4.6	6.4–8.3	6.5	9.8
Tensile strength psi×1,000		32–66	18.2–22.9	60–135			130–165	110	100–110
Specific gravity grams/cm³	19.32	8.94	10.5	8.9	9.34	12.44	8.25	7.73	8.75
Melting point, F	1945	1981	1761	2642	1125	3570		2700–2790	1885–1955
Modulus of elasticity psi×10⁶	12	17	10.3–11.8			42	30–31	29	19

Table II. Properties of Non Metals for High-Temperature Fuses

	Mycalex 2100[16]	Steatite Alsimag 35 L-3A[17]	Lime Glass Corning Code No. 0281[18]	Glass Fiber Electrical Tape ECCO 7 B[13]	Solder Glass Corning Pyroceram Cement No. 95 #2[14]
Electrical resistivity, ohm-cm	1.16×10^{16}	1×10^{14}–7×10^{2}	3.16×10^{8} (at 250 C)		
Thermal coefficient of expansion per F $\times10^{-6}$	8–8.7	4.83–5.13	4.83		5.0–6.1
Impact strength	0.6 ft-lb (Izod)	4.5 in-lb (Charpy)			
Water absorption, per cent	0	0–0.02 (impervious)			
Thermal stress resistance, F			32.4		
Thermal shock resistance, C			126		
Flexural strength, psi	13,000	18,000			
Compressive strength, psi	29,000	80,000			
Tensile strength, psi	6,000	8,500			
Specific gravity, grams/cm³	2.94	2.5	2.48–2.59	2.54	6.5
Softening temperature, F, and procedure used	1,000*	2,642 (ASTM† C24)	1100+ (ASTM C338)	1380–1550‡	840+¶
Dielectric strength, volts per mil	350	225 (1/4-inch specimen)			

* Maximum temperature maintained for 1,000 hours without sag on 1/8-inch×1/2-inch×3-inch specimen supported on 2-inch centers with no load.

† American Society for Testing Materials.

‡ Procedure not specified.

¶ Maximum continuous safe temperature.

usable for the fuse wires of the high-temperature fuse. However, a fuse wire will develop a higher temperature than its feeder wire. Tests have shown that silver and nickel are suitable materials even in an ambient temperature of 650 F and connected to feeder wires at 850 F. It would appear that an ambient temperature of 650 F will result in a decrease of about 20% in the nominal melting current. The presently used Pyrex glass housing has sufficient physical characteristics to be used in the new environment. However, other changes required in the fuse design, and discussed later on, made it advisable to adopt a different glass.

Significant physical characteristics of materials used in the development of high-temperature fuses are listed in Tables I and II.

As mentioned previously, larger current ratings require arc-suppressing fillers. Tests showed that woven glass fabric, and quartz particles, both of which are used in today's low-temperature fuses, are satisfactory at the new temperature level. Quartz fillers require a separate indicator element. Since this is not desirable, additional tests established that fiberglass plugs at the fuse wire ends in conjunction with a glass fiber liner would be equally effective in quenching the arc as quartz fillers.

The copper contact blades, but not their silver plating, were found to be satisfactory at higher temperatures. After considerable search for a plating that would withstand the new conditions, platinum and rhodium were found to be good replacements for the silver that failed through excessive oxidation and blistering. Rhodium was selected because it has better wear resistance and more stable contact resistance than platinum. An undercoat of nickel serves as a corrosion inhibitor.

One of the most serious problems in extending the temperature limit of present day fuses was the joining of the fuse housing to the end pieces that support the contact blades and act as finger grips. To solve this problem the configuration of the fuse had to be changed. It was obvious that the molded plastic caps were unsuitable at temperatures of approximately 900 F which is the temperature determined to be the highest reached in an ambient of 650 F for any part of the enclosure of a fuse connected to 850 F feeder wires. Since no other moldable organic material would do the job ceramic materials had to be substituted. Similarly, cements used for joining and sealing the glass housing to the caps would neither withstand the higher temperatures, nor, for that matter, would they bond glass to ceramics. Cements are also used in the low-temperature fuses to seal the contact blades where they pass through the end caps. This function they could not perform in the new environment.

It thus became necessary to:

1. Use a glass housing to accommodate the silver or nickel fuse wire; thus keeping it visible from the outside.
2. Employ ceramic end caps with finger grips.
3. Devise a means of joining the glass and the rhodium-plated contact blades to the ceramic parts.
4. Provide a seal as well as mechanical support.

After considerable experimentation a fuse design evolved that seemed to solve the problem. This design resulted from finding it quite feasible to join certain metals to ceramics by molding-in bushings, and then silver soldering the contact blades of the fuse wire assembly to these bushings. Also, certain glasses could be joined readily to ceramics by the use of solder glass. Fig. 8 shows the design. The fuse housing now became a ceramic, rectangular, open-top box with integral suitable finger grips and guards at the ends. Molded into this box were the two metal bushings to which the fuse contact blades in turn were silver soldered. The open top was then sealed off with a glass window. This design is simpler and stronger than two end caps joined to a glass housing. The question may be asked, "Why not mold the contact blades

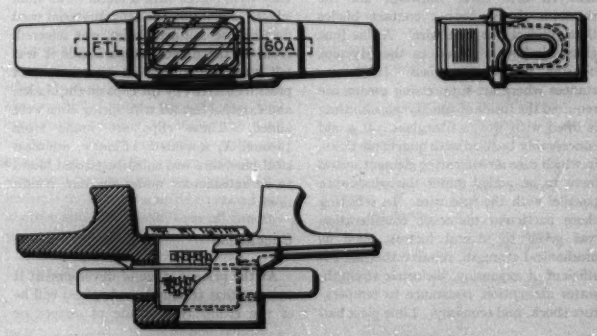


Fig. 8. High temperature, visually indicating fuse

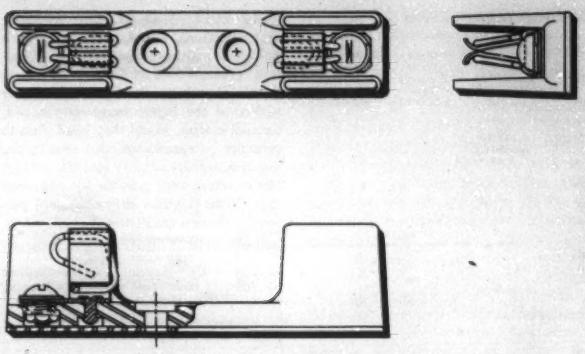


Fig. 9. High-temperature fuse holder

into the ceramic housing, eliminating the bushing?" For the contact blades a good electrical conductor is necessary. Copper and other good conductors have thermal coefficients of expansion incompatible with that of ceramics over so wide a temperature range. Stainless steel and other poor electrical conductors on the other hand have thermal coefficients of expansion that are sufficiently close to that of ceramics. By molding in bushings of the latter metals, and by silver soldering into these bushings the good electrical conductor needed for the contact blades, the difference of thermal expansion is now between metals rather than between metal and ceramic. The metals are sufficiently strong to withstand the resulting stress.

For the molded ceramic housing glass bonded mica in the form of Mycalex 2100 was selected. Molded into this material are stainless steel bushings, and silver soldered into these bushings are the rhodium plated copper contact blades that connect the fuse wire. A flat lime glass window is joined to the Mycalex housing, using solder glass. In the instances where arc-suppressing means are required the inside of the Mycalex housing is lined with woven fiberglass. It could conceivably be filled with quartz particles, in which case an indicating element would have to be added under the window in parallel with the fuse wire. In selecting these particular materials consideration was given to several factors, such as mechanical strength, relative thermal coefficient of expansion, dielectric strength, water absorption, resistance to temperature shock, and economy. Lime glass had to be substituted for Pyrex because so far it alone has been found to be solderable to the Mycalex housing. It withstands the required thermal shock.

Although the fuse holder presented fewer problems than the fuse, some changes had to be made to adapt it to the new environment. Fig. 9 illustrates the high-temperature design. The configuration of the low-temperature molded base was essentially adopted except that the material was changed from plastic to steatite, a ceramic that can be produced economically with the tolerances required, and which has the necessary mechanical and electrical properties. Changes also had to be made in the contact clips presently made from spring-tempered silver-plated beryllium copper, and in the silver-plated brass hardware used with present-day low-temperature fuse holders. The berryllium copper contact clips could only be used at the high temperature when they were plated with a metal other than silver. Rhodium, the same material used on the fuse contact blades, was selected. Also, at the higher temperatures it was felt necessary to provide greater contact pressure exerted by the clips on the blades, and external formed wire spring clips were added. These clips are made from Inconel *X*, unplated. Finally, stainless steel hardware was substituted and found to be satisfactory under the new conditions.

Conclusion

At the present stage of development it appears that terminals and splices will be of the crimp type, made of copper or nickel. Nickel terminals would be gold plated to obtain lower contact resistance. Copper terminals would be clad with stainless steel or plated with nickel and gold. It is not yet clear whether pre-insulated terminals and splices can be achieved by a combination of suitable crimping technique and material, or if the post insulated approach must be accepted. A solution of the problem in one direction or another seems likely.

The high-temperature fuse will in all probability be visually indicating, readily mounted in, and removed from its holder without the use of tools. It will embody the proven fuse wire and arc-suppressing means of present low-temperature fuses. However, it will be quite different from today's fuses in respect to the housing, which will be a one-piece, ceramic design.

Thus, the requirements presented by the substantially higher temperature of the environment could only be fulfilled by changing materials and design; mere substitution of materials by others that would withstand the higher temperatures was not sufficient.

References

1. Cable, Electric, Aircraft, High Temperature and Fire Resistant. *Military Specification MIL-C-25038*, Wright Air Development Center, Dayton, Ohio, Apr. 28, 1955.

2. Fuse, Aircraft, High Temperature. *Equipment Specification NAS-7255B*, North American Aviation, Inc., International Airport, Los Angeles Calif., July 9, 1957.

3. ASM, Metals Handbook. American Society for Metals, Cleveland, Ohio, 1948.

4. Copper-Base Alloy Mill Products. *Standards Manual*, Copper and Brass Research Association, New York, N. Y., Jan. 1958.

5. Nichrome and Other Nickel Electrical Alloys. *Catalog R56*, Driver-Harris Company, Harrison, N. J., 1956.

6. Expansion and Construction in Silver Alloy Brazing. *Technical Bulletin T-7*, Handy & Harmon, New York, N. Y., Sept. 1956; also Easy Flo 45, Characteristics and Properties, Apr. 1947.

7. Inco Nickel Alloy Helical Springs. *Technical Bulletin T-35*, International Nickel Company, Inc., New York, N. Y., May 1955.

8. ASME Handbook-Metals Properties, S. Hoyt, editor. McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

9. Wrought Alloys with Unique Properties. *Beryllium Copper Bulletin*, The Beryllium Corporation, Reading, Pa., 1956.

10. Glass Bonded Mica. *Mycalex Bulletin*, Haveg Industries, Inc., Taunton Division, Taunton, Mass. (Formerly General Electric Company)

11. *Alsimag Chart no. 573*, American Lava Corporation, Chattanooga, Tenn.

12. Properties of Selected Commercial Glasses. *Bulletin B93*, Corning Glass Works, Corning, N. Y., 1957.

13. The Inside Story of the New "Hercon Line" of Fiberglas Electrical Tapes. *Bulletin*, Horace Linton Division, Hess, Goldsmith & Company, Inc., New York, N. Y.

14. Properties-Pyroceram Brand Cement No. 95. *Product Information, P. P. I. No. 2*, Corning Glass Works, Apr. 2, 1958 (not available).

Load Characteristics of a Submerged-Arc Silicon-Smelting Furnace

GEORGE GRANT III
MEMBER AIEE

SUBMERGED-arc furnaces have been in commercial operation for more than 50 years. However, the literature tells little about what they require from the supplier of the electric energy or what effects they have on the supply system. A silicon-smelting furnace is typical of these operations. In fact, this process is generally considered to be one of the most exacting in its demands for precise operation to prevent undesirable performance of the furnace and unreasonable demands on the electricity supplier. This paper presents information on the electrical performance of such a furnace.

Equipment

Fig. 1 outlines the power circuit to the furnace. The utility company's supply at 46 kv is stepped down through a 10-mva (megavolt-ampere) transformer to 14.4 kv. It is then regulated by a step voltage regulator to 14.0 kv to the plant substation delta-connected bus. The circuit to the furnace building goes from this bus through a 500-mva air-magnetic circuit breaker and armored aerial cable. At the building, the auxiliary power transformer is connected directly to this cable while the 6.0-mva furnace transformer is connected through a disconnect and air-magnetic circuit breaker. Built into this transformer is added reactance, adjustable in five steps from the inherent 6.7% to 31.2%. The low voltage of the transformer is connected to interleaved bus bar. This bus bar is terminated and the delta connection closed at the flexible leads to the electrode clamps. Six electrode voltages may be selected at the operating panel through a remote control no-load tap changer in the transformer.

Control of the electrical load is attained by raising the electrodes to reduce the current and lowering them to increase the load. This is done with reversing electric motors driving hydraulic pumps which activate pistons that support the electrode with cables. Either manual or automatic control is available. The automatic control is piloted by a float-action current relay and each electrode is regulated independently.

Paper 58-1259, recommended by the AIEE Industrial and Commercial Power Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted June 25, 1958; made available for printing September 9, 1958.

GEORGE GRANT, III, is with the Dow Corning Corporation, Midland, Mich.

Process

Description

The process of smelting silicon consists of heating silica (silicon dioxide) with carbon. Carbon combines with the oxygen in the silica rock to form carbon monoxide and leaves the silicon in a molten state at the bottom of the furnace. The silicon is tapped from the furnace at regular intervals. Fresh mix or charge of silica and carbon is fed into the furnace from the top and banked around and between the electrodes. This is shown as *C* in Fig. 2. As the mix works its way down into the hot reaction zone toward the bottom, the carbon monoxide flows up through it. The carbon reaches the reaction zone to combine with the oxygen of the silica because it is protected from air by this blanket of carbon monoxide.

Conducting paths between electrodes are produced by the carbon in the charge. The conductivity of this path consists of a multitude of contacts and small arcs which continually change as the mix moves toward the bottom of the furnace and as the reduction process proceeds. Electric energy also goes from the electrodes through arcs to the metal in the bottom of the hearth. As a result, several factors affect the electric load in the furnace, with the multiple paths through the mix continually tending to make the load unsteady. By proper control of the mix and by the use of the proper voltage, the variations in conductivity through the multiple paths can be made to cancel one another and produce an over-all conductivity that is relatively steady.

Power Requirements

Ideal scheduling of operations calls for 24-hour operation at 6.0 mva. In practice, a monthly operating time factor of more than 90% is maintained. Most days do 97% or better. Maximum operating time is desirable both because of increased production and because delays upset the furnace.

When the power is shut off, the reaction stops and the reaction zone starts to cool. Carbon then starts to burn out of the charge because there is no protective carbon monoxide. The burning of the carbon disturbs the proportions of the mix and reduces the conductivity. Conductivity is further reduced by the cooling. This reduction in conductivity can become so great that the furnace cannot be restarted with the available voltages.

During the course of a month, shutdowns must be scheduled to make mechanical repairs. It has been found that both the duration of the shutdown and the condition of the furnace at the time of shutdown influence the ability to start up again. When the furnace was in good condition, it was started quite easily after a 48-hour outage. On the other hand, with the furnace in poor condition, an 8-hour delay has made restarting doubtful. Consequently, all scheduled shutdowns are aimed at less than 8 hours to minimize this chance of not getting started. Thus, it is difficult to state a maximum safe period for power discontinuity, but less than 8 hours can always be tolerated.

The process is not particularly sensitive to voltage variations but the electrical performance is. A 5% change causes changes in performance that are noticeable at once and become troublesome if maintained for more than 30 minutes. However, when a change of this nature arises, it is counteracted by changing taps on the transformer. A drop in the supply voltage is more desirable than an increase

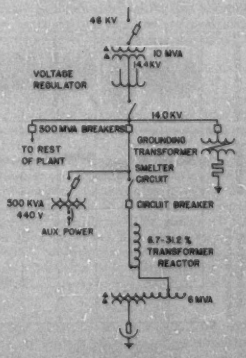


Fig. 1. Power circuit

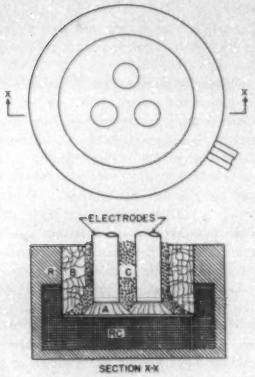


Fig. 2. Idealized section of furnace

A—Arc from electrode to metal
B—Burned-out mix, silica, and ash with no carbon
C—Fresh mix
R—Ceramic refractory
RC—Carbon refractory

because the tap change is made toward higher furnace voltage which results in a better power factor characteristic. However, voltage variations are not considered a serious problem.

Loads

General

The amount of load delivered for any given volt-ampere demand is determined by the circuit supplying power to the electrodes and the reactance of the furnace transformer is the major factor here. In Fig. 3 the megawatts and power factor

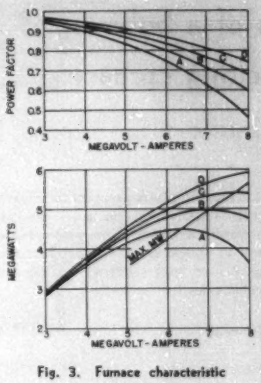


Fig. 3. Furnace characteristic

Transformer = 6.0 mva, 6.7% reactance
High voltage = 13.8 kv, 60 cycles per second
Low voltage = (A) 100, (B) 105, (C) 110, (D) 115

of the load are shown as functions of the megavolt-ampere demand. The relationship is shown for the four voltages found to work best in this installation. These show that it is best to operate at as high a voltage as possible because of the higher power and power factor for any specific volt-ampere demand. This also results in a steadier load. Further, it is seen that for each voltage there is a maximum power obtainable and further increases in volt-amperes result in reduced power and power factor. It is also apparent that as this maximum power is approached, there is little increase in power for substantial increases in demand with a corresponding serious drop in power factor.

In order to obtain the greatest production, the load is scheduled to be as high as possible within the limits determined by the electrical characteristics and thermal limitations of the electric equipment. A satisfactory compromise is 6.0 mva, which is the rating of the transformer and of the other electrical design features of the installation. Other loads are occasionally carried, but 6.0 mva is considered the normal load. The load is scheduled by megavolt-ampere demand, instead of megawatt load, in order to gain full utilization of the electric equipment and to minimize demand peaks, both of which are directly involved in changing voltage taps.

Measured Loads

Under normal operating conditions the load is steady with little fluctuation and no definite peak load time. Fig. 4(A) shows momentary and 15-minute load charts recorded when operation was considered to be good. (Momentary values are from thermal converter instruments having 2-second full-scale travel.) A period including a shutdown was included to show that the full 6.0-mva load is dropped and again picked up with no tapering off or gradual increase. The charts also show that no appreciable

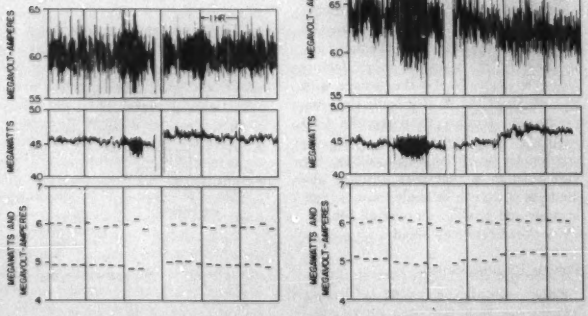


Fig. 4. Momentary and 15-minute load charts

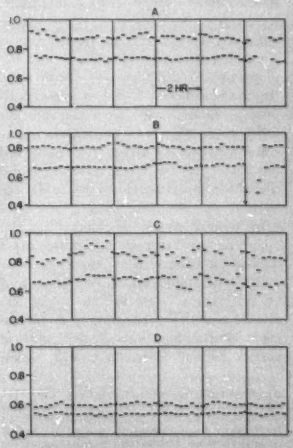


Fig. 5. Fifteen-minute load charts for several load conditions (7,200x); identifications are given in Table I

Table I. Identification of Charts in Fig. 5

Chart	Operation	Volts	Megawatts	Megavolt-Amperes	Power Factor
A	Good	110	5.2–5.6	6.2–6.8	0.80–0.87
B	Good	110	4.8–5.0	5.7–6.0	0.80–0.87
C	Bad	105	3.7–5.2	4.8–6.9	0.73–0.86
D	Good	100	3.8–4.0	4.2–4.5	0.87–0.92

inrush or surge is experienced from this type of switching. While operating, the power factor ranges from 0.80 to 0.87 and the demand has approximately 10% spread [per-cent spread = 100(maximum-minimum)/minimum]. In Fig. 5 and Table I, the 15-minute load charts *A*, *B*, and *D* demonstrate the same characteristics and spread in values for other loads. Under these conditions no flicker voltage problem has been noticed.

Unfortunately, these normal steady loads do not always exist and what might be called poor conditions sometimes present themselves. There are several process difficulties which produce trouble. Such conditions are usually corrected within 2 or 3 days. The momentary and 15-minute load charts in Fig. 4(B) show what happens under these conditions. In this case the problem only lasted 2 hours. Here the thermal demands remain substantially the same as for normal loads. However, the momentary values are distinctly worse with the spread increased to 30%. Slight flicker is sometimes produced. Infrequently, downright bad operation develops. No momentary charts are available to demonstrate this but a spread of more than 60% in demand is produced. Chart *C* of Fig. 5 and Table I shows that under these conditions the 15-minute demand spread goes up to 40% with corresponding large swings in power factor. This is accompanied by a severe flicker voltage problem. The ability of the operating personnel is a major factor in avoiding undesirable operating conditions, in minimizing the magnitude of the swings in load, and in shortening the duration when trouble is encountered.

Conclusions

This 6.0-mva load at 0.83 power factor is normally steady. Its prime requirement is to be able to go on and off the line without creating a disturbance on the power supply system. No surge is involved in direct switching, no flicker problem is present, moderate voltage fluctuations are compensated, and outages up to 8 hours can be tolerated. However, process difficulties may occasionally arise which result in heavy load and power factor swings and in a flicker voltage problem.

Discussion

H. Tauber (The Detroit Edison Company, Detroit, Mich.): The information that the author presents on the characteristics of a submerged-arc silicon-smelting furnace should be of great interest to those people designing power supplies for furnace loads.

The author points out that this load is normally stable, at good power factor, and generally free from flicker problems. He also points out that this submerged-arc furnace load tends to be unstable, but through proper control of the mix and the use of the proper voltage, the unstable tendencies can be cancelled out. He does, however, caution that there is the possibility of heavy load and power factor swings as well as a potential flicker problem.

I understand that not only is the submerged-arc furnace being used in silicon smelting, as described by the author, but also for phosphorous, calcium carbide, and ferroalloy production. Some of these furnaces are very large and even exceed the kilovolt-ampere capacity of the larger direct-arc furnaces. The nature of the charge and the size of the furnace may cause fluctuations that could be disturbing to the power supply.

We would expect that the submerged-arc furnace would present a high-power-factor resistance-type load that is relatively steady, because the electrodes are submerged into the charge. However, since I understand that part of the current is conducted through the charge by small arcs between the granular parts of the charge, I would expect that a fluctuating load is possible, depending on the properties of the charge. Although the author describes a fairly steady load for the silicon furnace, his words of caution concerning the possibility of fluctuation may become increasingly important when serving other type, or larger submerged-arc furnaces. For example, if the electrodes in a submerged-arc furnace are individually controlled, one would expect that, fairly frequently, one, two, or possibly all three of the electrodes would lose contact with the charge. The associated load fluctuation may be undesirable. In the case of the phosphorous furnace, "cave-ins" have been described by others that may cause a "splash" which changes the level of the charge or even a separation of the charge from an electrode. This, too, is a potential source of fluctuating load.

Unless the electric energy supplier is furnished with complete electrical load characteristics similar to that so carefully described by the author, I am of the opinion that the submerged-arc furnace should be considered as a potential fluctuating load, and should be served with the same caution the electrical supplier uses when serving direct-arc furnaces.

I should like to ask the author about this particular furnace and also about submerged-arc furnaces in general. Although a submerged-arc furnace is generally considered a resistance load under normal conditions, is the reactor in series with this furnace transformer used as a means of controlling load along with the taps on the transformer, or is it possibly used as a buffer to reduce voltage changes to the source? I would expect that, if arc stabilization is not a factor, minimum circuit reactance is desirable to give the best possible power factor. I would also like to know how often "bad" operation is experienced.

In an area where generally little information is available, the author should be complimented for supplying specific data on a submerged-arc furnace. I should like to ask if he has data available on other submerged-arc furnaces, and if he would expect that larger furnaces would undergo the same percentage change between "good" operation and "bad" operation, as does the furnace reported on?

A. M. Killin (Electro Metallurgical Company, Niagara Falls, N. Y.): Mr. Grant in his paper discusses the equipment and process for a 6,000-kva silicon-smelting furnace. The equipment consists of a standard submerged-arc ferroalloy furnace of suitable proportions for the manufacture of silicon metal. Smelting silicon consists of heating silica dioxide with carbon which produces the silicon metal with the release of carbon monoxide. A 6,000-kva transformer is provided for power source. This transformer is equipped with reactors adjustable in five steps from an inherent 6.7% to 31.2%. Controlled electric load of the furnace is obtained by raising and lowering the electrodes. Electrodes are controlled by a hydraulic system either manually or automatically controlled.

Operations of the silicon-smelting furnace are very steady. The monthly operating load factor is around 90%. Maximum operating time is beneficial not only in increased production but in the reaction in the furnace.

In Mr. Grant's paper he has brought out variances and operating conditions due to load changes. In some cases it may be 2 to 3 days in order to correct poor furnace operating conditions. The ability of the operating personnel is a major factor in avoiding undesirable operating conditions and minimizing the magnitude of load swings.

George Grant: I am not associated with any installation other than that presented in the paper and am therefore in no position to discuss such equipment.

Mr. Tauber points out the fact that submerged-arc furnaces are used for phosphorous, calcium carbide, and ferroalloy production. There are many more of these than there are silicon-smelting operations, and the manufacture of some refractories and abrasives can be added to the list. Some of these installations use as much as 35,000 kva.

Individually controlled electrodes can theoretically produce conditions such as suggested by Mr. Tauber. However, in our installation using this type of control, no problem has been experienced that was traceable to such conditions.

The transformer was originally built for an entirely different application. The series reactor was included to fit that job. In this operation it is not used for arc stabilization. Reactance of less than the inherent 6.7% woud give more desirable power, power factor, and voltage regulation characteristics. Added reactance has been tried as a buffer to reduce voltage changes to the source. This produced dubious results.

"Bad" operation has been experienced three times since the furnace was put into service in 1954. These were all encountered during the first year of production while entirely inexperienced personnel were learning the rudiments of such an operation. Generally speaking, bad and poor load demands are produced when more load is impressed on the furnace than its condition will gracefully accept.

It might be expected that bigger furnaces operated poorly will result in somewhat the same percentage change between good and bad load swings.

As pointed out in the paper, there is no question that in smelting silicon the steadiness of the load is primarily a function of the ability and philosophy of the operating personnel.

Electrified Fibrous Air Filters

J. W. THOMAS
NONMEMBER AIEE

E. J. WOODFIN
NONMEMBER AIEE

THERE are two principal methods of separating suspended particles from air: mechanical separation and electrostatic precipitation. Mechanical separation includes dry-type filters, cyclones, water scrubbers, and gravity settling chambers. Dry-type filters are commonly used in air conditioners. Mechanical separators depend on inherent differences in size and density existing between particles and the gas in which the particles are suspended. In contrast, electrostatic precipitators cause a difference between properties of particle and gaseous medium; that is, either the particle is charged in an ionizer[1] or, without an ionizer, the particle undergoes polarization in the field existing between charged plates of the precipitator.

The principal mechanism of removal in mechanical separators of preponderant interest is inertial impingement. In essence, this removal method consists of turning the direction of the air stream, which causes particles to be thrown out by their inertia onto collecting surfaces. Unfortunately, it is necessary to exert force on the entire air stream. Hence, the operation causes a pressure drop which increases monotonically with separation efficiency.

Electrostatic precipitators do not suffer from this defect. It is not necessary to turn the direction of the air stream since the applied field acts directly on the particles alone. Hence electrostatic precipitators have the inherent advantage of high efficiency at very low pressure drop. However, ordinary electrostatic precipitators are not satisfactory where size is at a premium. Size of the precipitator can be reduced without sacrificing efficiency by spacing the plates closer together, but there are definite limits to this process.

However, choice is not limited to pure mechanical filters or electrostatic precipitators. It is possible to insert Fiberglas into an electrostatic precipitator, giving an "electrified fibrous air filter." The principal feature of this combination filter may be seen from the following example. Consider a parallel plate precipitator having a plate spacing of 1 inch, with a particle-laden air stream moving through parallel to the plates. Because of the field gradient between the plates a charged particle will have a velocity component perpendicular to the air flow direction. A particle in the center of the spacing will have to travel 1/2 inch perpendicular to the air flow direction before being deposited on a surface. Consequently, if air velocity is too high or the precipitator too short in the direction of air flow, the particle will not be collected. However, if the spacing is now packed with a low-density fibrous filtering material, the particle will have only a few thousandths of an inch to travel under influence of the field before contacting a collecting surface. Thus much higher efficiency can be expected at some sacrifice in pressure drop.

This investigation of electrified fibrous filters was undertaken because of deficiencies of pure mechanical filters or electrostatic precipitators for air-conditioner application. Principally, it was desired to obtain filters that were relatively small and cheap yet which had low-pressure drops and high efficiency in removing cigarette smoke. Efficiencies of electrified fibrous filters were determined without attempting to precharge the test aerosol, using cigarette smoke as a test aerosol. This is a common smoke, and is made conveniently by a cigarette-smoking machine. Average size of cigarette smoke is below 1 micron in diameter probably about 1/2 micron, although authorities differ.[2] Cigarette smoke is much more difficult to filter mechanically

Paper 59-680, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Appliance Technical Conference, Cleveland, Ohio, May 18-19, 1959. Manuscript submitted September 22, 1958; made available for printing September 10, 1959.

J. W. THOMAS and E. J. WOODFIN are with the Whirlpool Corporation, St. Joseph, Mich.

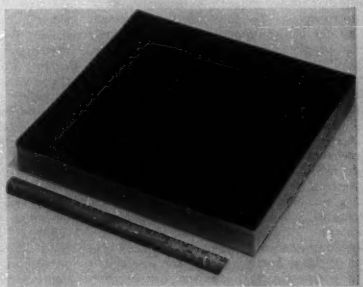


Fig. 1. Electrified pleated filter

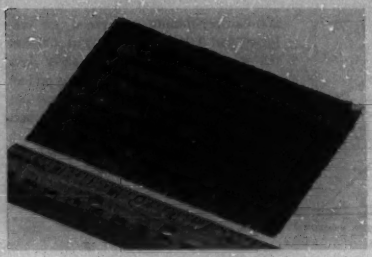


Fig. 2 (left). Section of filter of Fig. 1

Fig. 3 (right). One-inch precipitator

than larger particles such as ragweed pollen, and its is probable that if a filtering device efficiently removes cigarette smoke it will also efficiently remove any other air-borne particulate matter.

Test Filters and Precipitators

Fig. 1 shows a commercial-type electrified fibrous filter. This filter is of Fiberglas media, electrified, and pleated to provide a very large surface area which greatly reduces the pressure drop through the filter. This particular filter has 8 square feet of filtering area per square foot of superficial area, which is the area enclosed by the outside edges of the filter. Fig. 2 shows the unfolded Fiberglas media. The dark lines spaced 0.31 inch apart, are conducting channels. Alternate channels are at a 3,000-volt potential with respect to adjacent channels which are grounded.

Two test electrostatic precipitators were made from a 4- by 4-inch Lucite housing, slotted at top and bottom to receive 0.060-inch-thick aluminum or copper plates. One precipitator was 14.5 inches long in the direction of air flow, and used plates 4 inches wide by 14 inches long spaced 0.91 inch apart. The other precipitator was similar but was only 1 inch long in the direction of air flow. Fig. 3 shows the 1-inch precipitator. Both precipitators were used empty, and packed with Fiberglas.

Test Procedure

Filters and precipitators were tested in ducts at 50-, 200-, and 400-foot-per-minute superficial face velocity. Concentrations of cigarette smoke influent and effluent from a test filter were determined by sampling about 0.3 cubic foot of smoke through a high-efficiency white filter paper. It was not necessary to sample isokinetically since the particle size of cigarette smoke is sufficiently small that errors do not arise from nonisokinetic conditions.

Relative amounts of cigarette smoke deposited on the white filters were determined by measuring the reflectance of the brown stains caused by the cigarette smoke. From reflectance readings, efficiency of the test filter was determined.

It was necessary to convert the reflectance reading of each sampling filter into a quantity proportional to mass of cigarette smoke on the filter. Since the Kubelka-Munk equation has been used in cloth-soiling studies[3] to convert reflectance readings into quantities proportional to mass of soil of the cloth, it is possible that this equation also could be used for assessing cigarette smoke. The equation is

$$K/S=(1-R)^2/2R$$

Quantity K/S is the ratio of coefficient of reflectivity to coefficient of light scattering. Under certain conditions this quantity is a linear function of mass of discoloring component present on the surface of a white object. Quantity R is fractional reflectance of the sample referred to reflectance of magnesium oxide as unity. If the equation holds for cigarette smoke a plot of $(K/S)_s-(K/S)_o$ against quantity of smoke collected should give a straight line with zero intercepts. Subscript s refers to the sample filter and subscript o to a clean white filter. Two experiments were made to test this relationship.

In each experiment a cloud of cigarette smoke was set up in a 1,200-cubic-foot closed room. This smoke dissipates very slowly. Multiple smoke samples were taken for known, different times, giving known ratios of smoke quantity on the sampling filters. Fig. 4 shows the results. The two curves are at different smoke concentrations. Both curves show that the Kubelka-Munk equation applies fairly well although there is some divergence from proportionality.

Test filter efficiencies were therefore calculated as follows where the subscript A refers to the influent sample, subscript B to the effluent sample, and subscript o to a clean white filter.

$$\text{Per-cent efficiency}=\frac{(K/S)_A-(K/S)_B}{(K/S)_A-(K/S)_o}$$

Results of tests of the pleated filter,

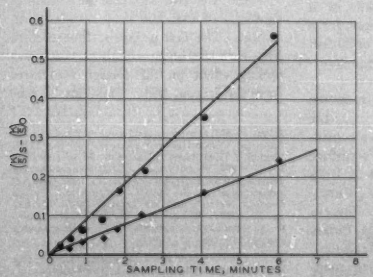


Fig. 4 (left). Test of Kubelka-Munk equation

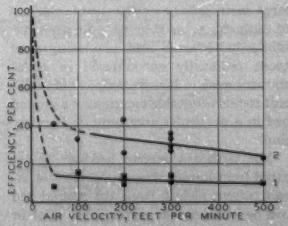


Fig. 5 (right). Efficiency of pleated filter

1—No field
2—9,600-volt-per-inch field

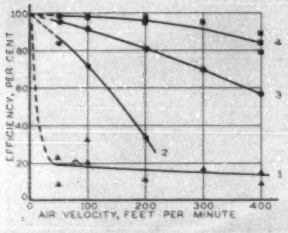


Fig. 6. Efficiency of 14.5-inch precipitator

1—Fiberglas, no field
2—No Fiberglas, 23,500-volt-per-inch field
3—Fiberglas, 10,700-volt-per-inch field
4—Fiberglas, 23,500-volt-per-inch field

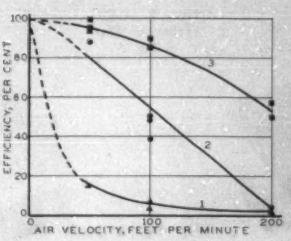


Fig. 7. Efficiency of precipitator of Fig. 3

1—5,500-volt-per-inch field
2—11,000-volt-per-inch field
3—20,000-volt-per-inch field

illustrated in Figs. 1 and 2, are given in Fig. 5. Pressure drop was 0.02 inch of water at 200-foot-per-minute superficial face velocity, and 0.04 inch of water at 400 feet per minute. The curves show that application of the field more than doubled efficiency at flow rates between 50 and 500 feet per minute.

Fig. 6 shows performance data of the 14.5-inch precipitator. Pressure drop when Fiberglas loaded was 0.18 inch of water at 200-foot-per-minute superficial air velocity and 0.55 inch of water at 400 feet per minute. Without Fiberglas, pressure drop was less than 0.01 inch of water at both velocities. Curves *3* and *4* show improved efficiency resulting from impressing a field gradient on Fiberglas packing. Improvement was better than a factor of five, as shown by a comparison of curves *1* and *4*, in the air velocity range of 50–400 feet per minute.

Also, results show that for a precipitator operating on natural cigarette smoke a great improvement in performance may be obtained by packing the precipitator with Fiberglas, at some sacrifice in pressure drop.

Fig. 7 shows performance data of the 1-inch precipitator illustrated in Fig. 3. Pressure drop in this Fiberglas-packed precipitator was 0.03 inch of water at 200-foot-per-minute superficial air velocity and 0.06 inch of water at 400 feet per minute. Curves *1* to *3* show data taken with both Fiberglas packing and an impressed electrical field. Data were not obtained for electrical field only or Fiberglas only as efficiencies were too low to measure accurately, being of the order of 2% or less. Again, the data show great improvement caused by impressing an electrical field on Fiberglas material.

Conclusions

1. Cigarette smoke filtration efficiency of Fiberglas filtering media can be improved greatly by impressing a field gradient on the media.

2. Improvement is synergistic: that is, efficiency of a Fiberglas-packed precipitator is considerably greater than that calculated from efficiency of each effect separately.

3. Air-conditioner filters 1 inch thick can be made to have 50% efficiency in filtering cigarette smoke at 200 feet per minute at less than 0.05-inch pressure drop by using Fiberglas-packed electrostatic precipitators.

References

1. Particle Charging in Electrostatic Precipitation H. J. White. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1186–91.

2. Concentration and Particle Size of Cigarette Smoke Particles, Gerhard Langer, M. A. Fisher. *Archives of Industrial Health*, Chicago, Ill., vol. 13, 1956, pp. 372–78.

3. Detergent Action, O. C. Bacon, J. E. Smith. *Industrial and Engineering Chemistry*, Washington, D. C., vol. 40, 1948, pp. 2361–70.

Deep Ground Beds for Cathodic Protection

S. E. TROUARD
MEMBER AIEE

E. A. WAGNER, JR.
MEMBER AIEE

THE SUCCESSFUL application of cathodic protection to undergound structures in the crowded confines of a large metropolitan city presents problems of cathodic protection co-ordination certainly more acute and perplexing than those ordinarily encountered in cross-country work. It is frequently very difficult to find a suitable location for a ground bed in a particular area requiring protection.

The farther a ground bed can be located from foreign structures, the less direct current pickup from the ground bed will be found on the structures. Yet in a large city it may be impossible, or at least impracticable, to find a spot for a ground bed where more than 20 or 30 feet horizontal separation from foreign structures can be secured.

The radical new concept of introducing cathodic protection currents into the earth from a ground bed located a considerable distance below the earth's surface was originally suggested in 1951 by Robert J. Kuhn, New Orleans, La., Corrosion Consultant. As a result of the co-operative effort of New Orleans Public Service Inc., engineers and of Mr. Kuhn, the designs for deep rail ground beds mentioned in this paper were developed. The first deep rail ground bed in New Orleans was installed on December 21, 1952. This new idea of deep ground beds has received rather wide public acceptance as evidenced by the fact that a large number of deep ground beds have since been installed by other utilities, particularly in the southern states, and even as far away as Iran.

There are two separate and distinct problems of cathodic protection co-ordination. The first problem, illustrated in Fig. 1, is the anode effect, or the direct pickup effect on the foreign structures from the ground bed. This can be ren-

Paper 59-505, recommended by the AIEE Chemical Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southeast and South Central District Meeting, Atlanta, Ga., April 8–10, 1959, and re-presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted January 8, 1959; made available for printing May 22, 1959.

S. E. Trouard and E. A. Wagner, Jr., are with New Orleans Public Service Inc., New Orleans, La.

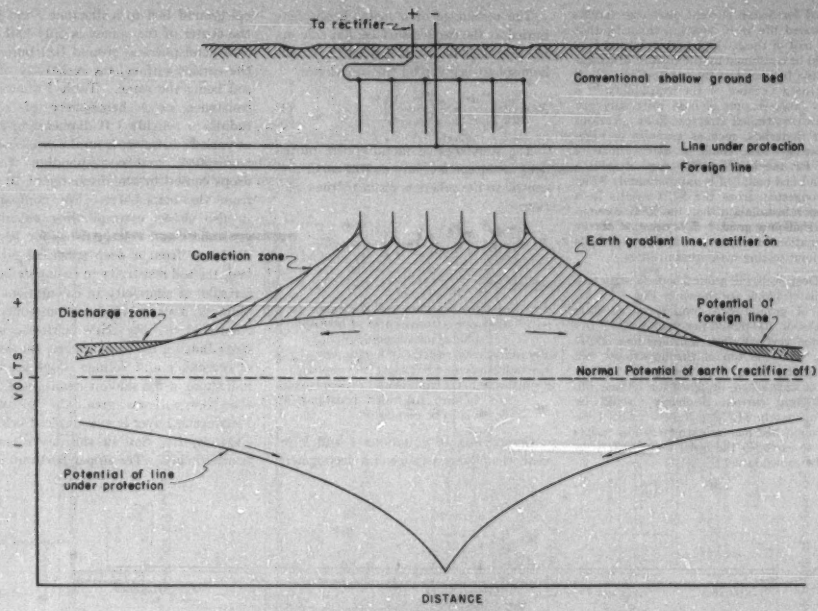


Fig. 1. Direct ground bed effect

Schematic potential relationships of earth, foreign line, and line under protection from conventional shallow ground bed

dered so small as to be undetectable by placing the ground bed at a great distance from the foreign lines. With conventional shallow ground beds this would be most difficult to accomplish where the foreign lines are in the form of a network with mains on every street and cross street.

The second problem may be called the cathode effect. This effect is independent of the location of the ground bed. It is beyond the scope of this paper to discuss ways and means of coping with this particular phase of the co-ordination problem.

For the cathodic protection of the coated steel gas main system in New Orleans, three different types of ground beds are employed:

1. Conventional shallow rail ground bed as shown in Fig. 2. Provided a suitable location can be secured, this type of ground bed is generally preferable because it is the least expensive ($250 installed). The used rails weigh approximately one ton and such a ground bed is designed conservatively for a 10-year life with a 5-amp (ampere) current discharge, assuming that only half of the metal would go into solution, and the rest lost by segregation. Actually, in this design the inner faces of the rail webs will discharge but little current because of mutual interference effects. Thus, rather than obtain only 50% use of the metal, closer to 75% would be secured and the losses from segregation would be reduced, hence the actual life of the ground bed would be longer than the 10-year figure mentioned. In 500-ohm-cm (centimeter) soil, the resistance to earth of this type of bed ranges from 0.5 to 1.0 ohm.

2. Deep rail ground bed, construction details of which are shown in Fig. 3. In this type of ground bed, a hole is drilled in the earth nominally 8 inches in diameter, but actually closer to 9 inches in diameter, 250 to 300 ft (feet) in depth. Two 30-ft lengths of 100-pound-per-yard used T-rails are lowered into the hole by means of a 1/4-inch-diameter wire rope. The rails are joined together longitudinally by means of rail joint plates, and a separate wire with no splices is brought from each rail length to the surface of the earth. The resistance to earth of this type of ground bed is in the general order of 0.2 ohm, excluding lead-wire resistance. The installation is de-

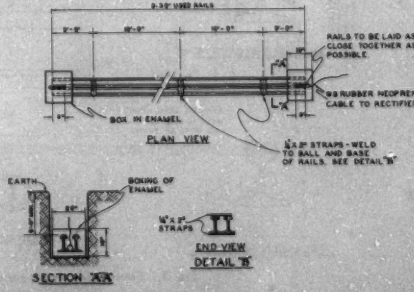


Fig. 2. Typical rail ground bed for cathodic protection

signed for 5-amp current discharge and its estimated life is 10 years, estimating that only half of the 2,000-pound weight of iron would be consumed by the current discharge and the balance lost by segregation. Rails were used because of the availability of a fairly large surplus of used rails, salvaged from discontinued streetcar lines. Various other materials, such as graphite or high-silicon cast iron, could be used instead of rails for the anode. The cost of such a ground bed installed is approximately $750. In congested areas the $750 results in a cheaper installation than the $250 conventional shallow ground bed because of the elimination of expense of trial ground beds and joint testing with other utilities.

3. Deep stalk rail ground bed, construction details of which are shown in Fig. 4. This type of ground bed consists of 12–30-ft lengths of 100-pound-per-yard used T-rail inserted vertically into a drilled hole 360 ft deep, with the top of the uppermost rail practically flush with the ground. The cost of such a bed, designed for 10-year life at 30-amp current discharge, would be approximately $1,750 installed. This type of ground bed has an extremely low resistance to earth, in the order of 0.05 ohm, as will be shown later.

The resistance of a metal hemisphere buried at the earth's surface, flat side up, to any shell of earth concentric to the hemisphere is given by the equation:

$$Rr = \frac{\rho}{2\pi}\int_{r_1}^{r}\frac{dr}{r^2} = \frac{\rho}{2\pi}\left(\frac{1}{r_1}-\frac{1}{r}\right) \qquad (1)$$

The resistance of metal sphere buried deep in the earth to any shell of earth concentric to the sphere is given by the equation:

$$Rr = \frac{\rho}{4\pi}\int_{r_1}^{r}\frac{dr}{r^2} = \frac{\rho}{4\pi}\left(\frac{1}{r_1}-\frac{1}{r}\right) \qquad (2)$$

where

Rr = resistance in ohms of the earth to a distance r from center of hemisphere (or sphere) measured in cm

ρ = uniform soil resistivity in ohm-cm

r_1 = radius in cm of hemisphere (or sphere)

r = distance in cm from center of hemisphere (or sphere) to which resistance to earth is to be computed

Comparison of equations 1 and 2 reveals that the resistance of a deep spherical ground bed to a distance r cm from the center of the sphere is only half that of a hemispherical ground bed buried at the earth's surface, the resistivity of the soil being the same. Table I shows the resistance of a hemisphere of 15-cm radius, or roughly 1 ft diameter to earth at various distances from the center of the hemisphere, and corresponding voltage drops caused by the discharge of 10 amp from the hemisphere. For comparison it also shows corresponding calculated resistances and voltage drops for 10-amp discharge from a deep spherical ground bed, the soil resistivity in each case being, for sake of simplicity in calculations 628 ohm-cm, a value which is commonly encountered in the New Orleans area. Note that for the deep sphere, the values of resistance and voltage drops are only half those of the shallow hemisphere. In the New Orleans area, the so called Pleistocene Layer is encountered below a 75-ft depth. Soil in this layer is fine textured clay. The upper portion of the

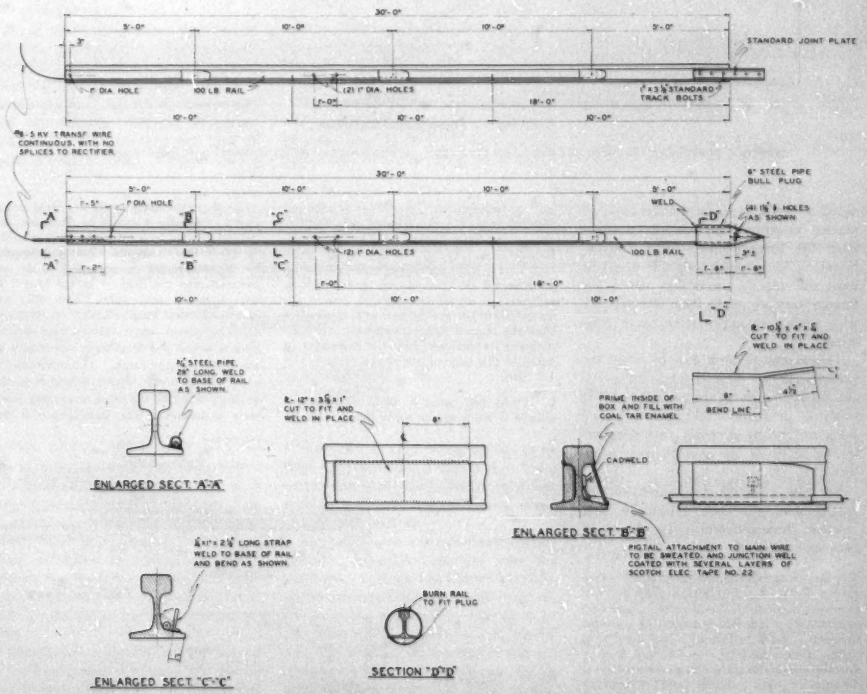


Fig. 3. Construction details for standard deep vertical rail ground bed

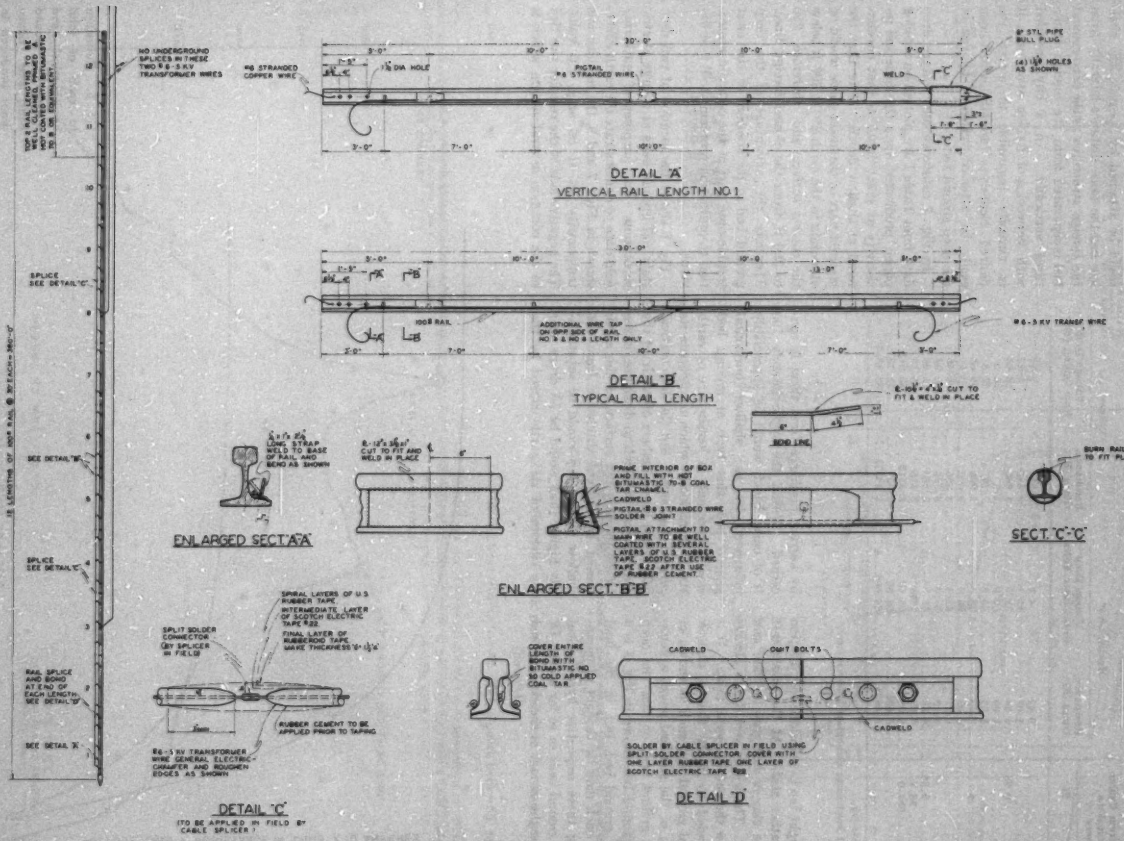

DETAIL 'A'
VERTICAL RAIL LENGTH NO.1
DETAIL 'B'
TYPICAL RAIL LENGTH
ENLARGED SECT. A-A
ENLARGED SECT. B-B
SECT. C-C
DETAIL 'C'
DETAIL 'D'
NO UNDERGROUND SPLICES IN THESE TWO #8 - 5 KV TRANSFORMER WIRES
#8 STRANDED COPPER WIRE
PIGTAIL #8 STRANDED WIRE
WELD
6" STL. PIPE BULL PLUG
100 LB RAIL
ADDITIONAL WIRE TAP ON OPP. SIDE OF RAIL NO. 9 & NO. 8 LENGTH ONLY
#6-5 KV TRANSF. WIRE
PRIME INTERIOR OF BOX AND FILL WITH HOT BITUMASTIC TO #8 COAL TAR CHANNEL
CADWELD
PIGTAIL #8 STRANDED WIRE SOLDER JOINT
PIGTAIL ATTACHMENT TO MAIN WIRE TO BE WELL COATED WITH SEVERAL LAYERS OF U.S. RUBBER TAPE & SCOTCH ELECTRIC TAPE #33 AFTER USE OF RUBBER CEMENT
BURN RAIL TO FIT PLUG
SPIRAL LAYERS OF U.S. RUBBER TAPE
INTERMEDIATE LAYER OF SCOTCH ELECTRIC TAPE #33
FINAL LAYER OF RUBBERIZING TAPE WIRE THICKNESS 0-1/16"
SPLIT SOLDER CONNECTOR BY SPLICER IN FIELD
RUBBER CEMENT TO BE APPLIED PRIOR TO TAPING
#6-5 KV TRANSFORMER WIRE GENERAL ELECTRIC CHROMIUM EDGES AS SHOWN
COVER ENTIRE LENGTH OF BOND WITH BITUMASTIC NO 50 COLD APPLIED COAL TAR
CADWELL
CUT BOLTS
SOLDER BY CABLE SPLICER IN FIELD USING SPLIT SOLDER CONNECTOR COVER WITH ONE LAYER RUBBER TAPE ONE LAYER OF SCOTCH ELECTRIC TAPE #33
SPLICE SEE DETAIL 'C'
RAIL SPLICE AND BOND AT END OF EACH LENGTH SEE DETAIL 'D'
SEE DETAIL 'A'
SEE DETAIL 'B'

Fig. 4. Construction details for 360-ft length of vertical rail ground bed

Table I. Voltage Values in Earth Surrounding a Hemispherical Ground Bed at Earth Surface and a Deep Spherical Ground Bed, Each Discharging 10 Amp

Distance From Center of Hemisphere or Sphere (r)		Hemisphere*		Sphere*	
Cm	Ft	Resistance, Ohms (Rr)	Voltage Drop from Center, Volts	Resistance, Ohms (Rr)	Voltage Drop from Center, Volts
15		0.42	4.2	0.21	2.10
20		1.67	16.7	0.885	8.85
30	0.98	3.33	33.3	1.665	16.65
60		5.00	50.0	2.5	25.0
90		5.56	55.6	2.78	27.8
120		5.83	58.3	2.965	29.65
150		6.00	60.0	3.000	30.0
180		6.11	61.1	3.055	30.55
210		6.19	61.9	3.095	30.95
300	9.84	6.33	63.3	3.165	31.65
3,000	98.4	6.63	66.3	3.315	33.15
30,000	984.0	6.663	66.63	3.3315	33.315
300,000		6.666	66.66	3.333	33.33
3,000,000		6.6666	66.666	3.33333	33.3333

ρ = 628 ohm-cm.
* Radius of hemisphere and sphere is 15 cm (r_1).

layer is light tan in color due to oxidation and is rather stiff. At the 250–300-ft depth the clay is dull gray in color and less stiff. The soil is wet, being always well below the natural water table. It is not intended to imply that a shallow horizontal ground bed such as shown in Fig. 2 is, in the strictest sense, a hemispherical ground bed, or that the deep vertical ground bed shown in Fig. 3 is a spherical ground bed, but, in the final analysis, each may be thought of as approaching these respective hypothetical types of ground beds in so far as the *IR* drops near the surface of the earth caused by current discharge from them is concerned. Within the approximately 100 ft of earth which would be appreciably influenced by the shallow horizontal ground bed, this type of ground bed would have only a hemisphere of earth 100 ft in radius in which to dissipate its current discharge. But the deep vertical bed would have a sphere of earth, of the same radius, in which to dissipate, hence *IR* drops near the surface of the earth caused by current discharge from the shallow ground bed would be substantially higher than would be *IR* drops near the surface of the earth caused by the same amount of current discharging from the deep vertical ground bed. For all practical purposes, the same net effect could be attained by locating the shallow bed the same distance laterally from the foreign structures as the deep ground bed is separated from the foreign structures vertically. In a city crowded with networks of foreign structures, it is virtually impossible to secure the desired clearances laterally which can easily be attained vertically by use of the deep ground bed.

In Fig. 5 resistance and voltage drop data from Table I are plotted on semilog paper, predicated upon the same soil resistivity at great depth as that encountered at shallow depth. Note that the steepest part of the lower curve would be at points in the earth 250 to 300 ft below the earth's surface and, therefore, 250 to 300 ft from the closest foreign lines. Therefore the advantage of the deep ground bed in the solution of the most troublesome part of the cathodic protection co-ordination problem should be obvious.

Fig. 6 shows the location of two deep stalk rail ground beds at a large electric substation. For purposes of gathering field data to demonstrate the earth effects shown in Figs. 7 and 8, the rectifier was temporarily set at 55 amp, 7.0 volts. The negative side of the rectifier was tied to the common neutral system which is

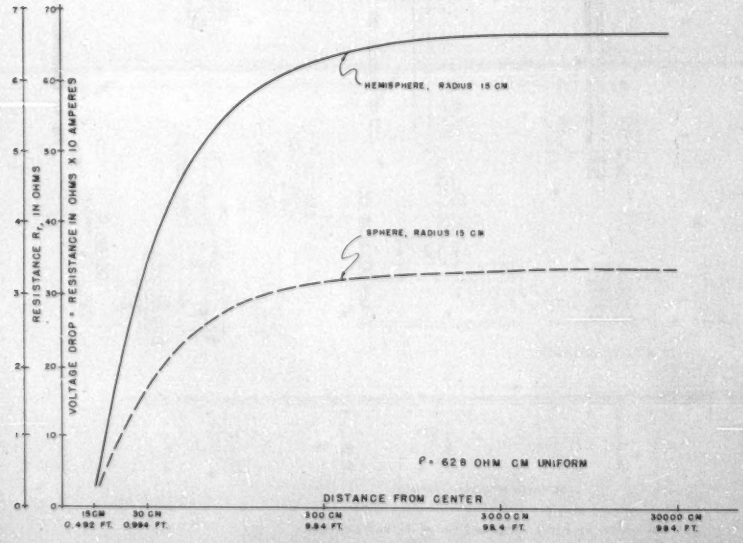

HEMISPHERE, RADIUS 15 CM
SPHERE, RADIUS 15 CM
RESISTANCE R_r IN OHMS
VOLTAGE DROP = RESISTANCE IN OHMS X 10 AMPERES
ρ = 628 OHM CM UNIFORM
DISTANCE FROM CENTER
15CM 0.492 FT.
30 CM 0.984 FT.
300 CM 9.84 FT.
3000 CM 98.4 FT.
30000 CM 984. FT.

Fig. 5. Resistance to earth of solid hemispherical and spherical ground beds, and voltage drops produced in earth by each with 10-amp discharge

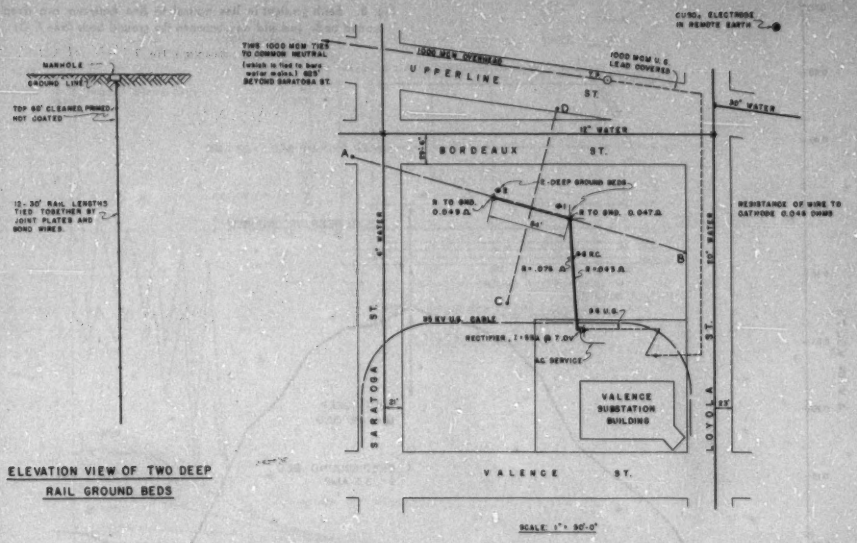
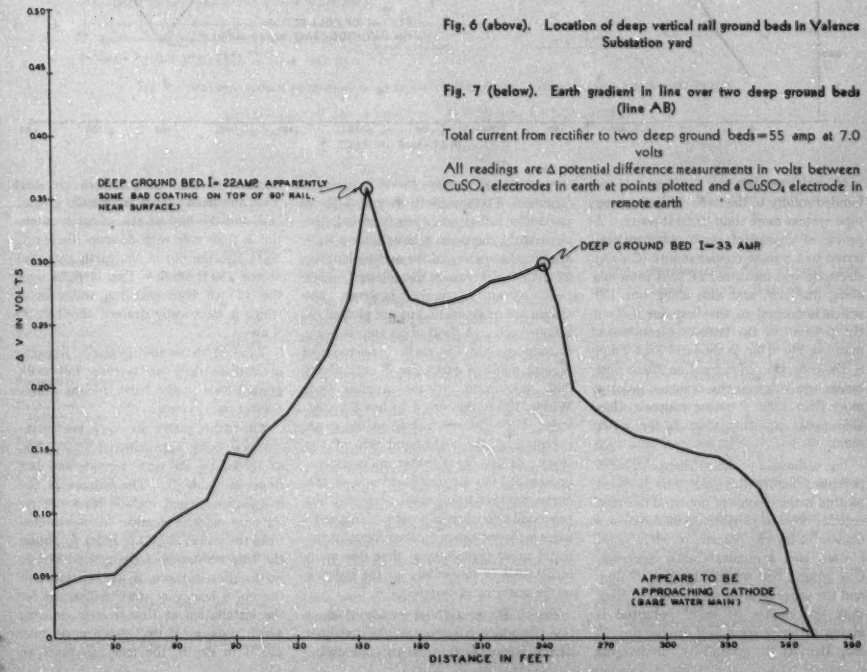


Fig. 6 (above). Location of deep vertical rail ground beds in Valence Substation yard

Fig. 7 (below). Earth gradient in line over two deep ground beds (line AB)

Total current from rectifier to two deep ground beds = 55 amp at 7.0 volts

All readings are Δ potential difference measurements in volts between $CuSO_4$ electrodes in earth at points plotted and a $CuSO_4$ electrode in remote earth



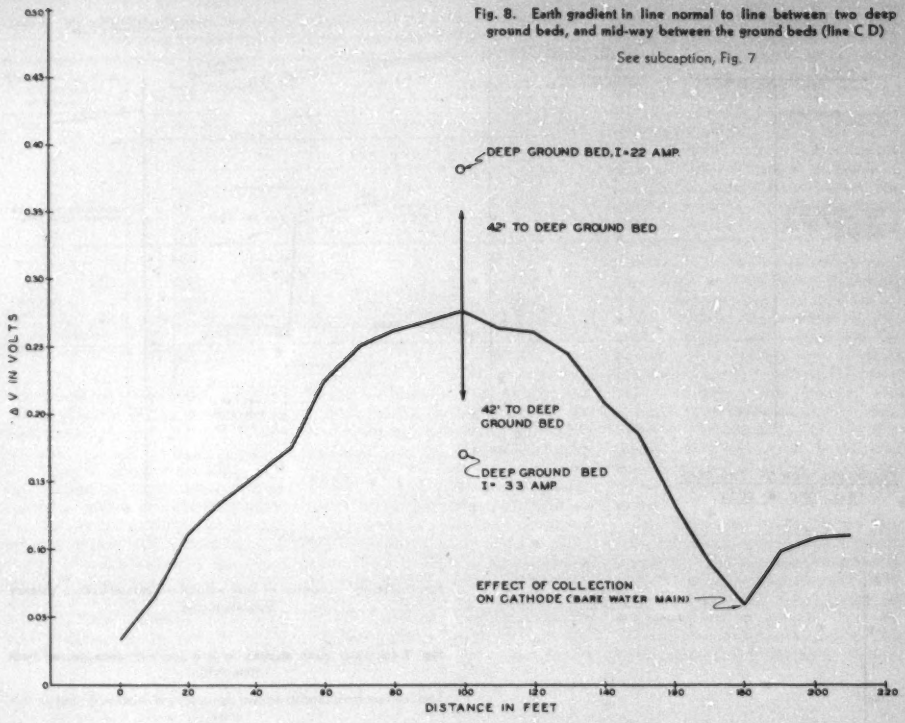


Fig. 8. Earth gradient in line normal to line between two deep ground beds, and mid-way between the ground beds (line C D)

See subcaption, Fig. 7

bonded solidly to the bare cast-iron water pipe system more than 1,000 ft away. A profile of change of earth potentials referred to a remote copper sulfate ($CuSO_4$) electrode was measured at 10-ft intervals along line *AB*, and also along line *CD* which is normal to the first line. Note the position of the bare cast-iron water mains in Fig. 6 for, as shown in Figs. 7 and 8, the collection of current on these bare mains, even though the drainage point is more than 1,000 ft away, causes noticeable earth gradients towards the water mains.

Fig. 7 shows a profile of change of earth potential across line *AB*. Each point plotted is the change of potential between rectifier-off and rectifier-on readings of a $CuSO_4$ electrode placed at each point plotted and a remote $CuSO_4$ electrode. One ground bed was discharging 22 amp and the other 33 amp, at 7.0 volts. Note that the change of earth potential is higher at the ground bed discharging 22 amp than at the ground bed discharging 33 amp. The top 60 ft of each of these particular rail ground beds is coated, but apparently there must have been a bare spot on the coating of the one discharging 22 amp which causes the relatively high peak shown. Actually, however, the values are quite small, but are plotted on a large scale. A total of 55 amp is being introduced into the earth here and the highest point of earth rise is only about 0.35 volt right at the ground bed. Within 100 ft this drops to less than 0.1 volt. Note the sharp drop in the earth potential at the right-hand side of the curve, as the bare water main is approached. On the left-hand end of the curve, another water main is crossed but practically no effect is shown. This indicates high-resistance caulked joints in the water main system, such that this main is not tied in too solidly to the balance of the water main system.

Fig. 8 shows another profile along a line *CD* which is normal to the first line, and mid-way between the two rail stalks. This line passes 42 ft from each ground bed, and the highest rise of earth potential is 0.27 volt with 55-amp discharge. Note also the dip in the earth potential at the 180-ft mark. This is right over the 12-inch bare cast-iron water main which is electrically drained about 1,000 ft away.

Table II shows the resistance to earth of various deep or vertical rail stalk ground beds in the New Orleans Public Service Inc., system.

The 300-ft stalks are very low in resistance, being approximately 0.05 ohm, as shown by the first, second, and last items in Table II. The balance of the installations listed include lead-wire resistance which should be subtracted from the values shown in order to obtain the true resistance of the ground bed to earth. For instance, if the resistance of the no. 8 lead wire were subtracted for the installation at Roman near Tulane, which consists of two 30-ft-length rails, 250 ft in depth, the true resistance to

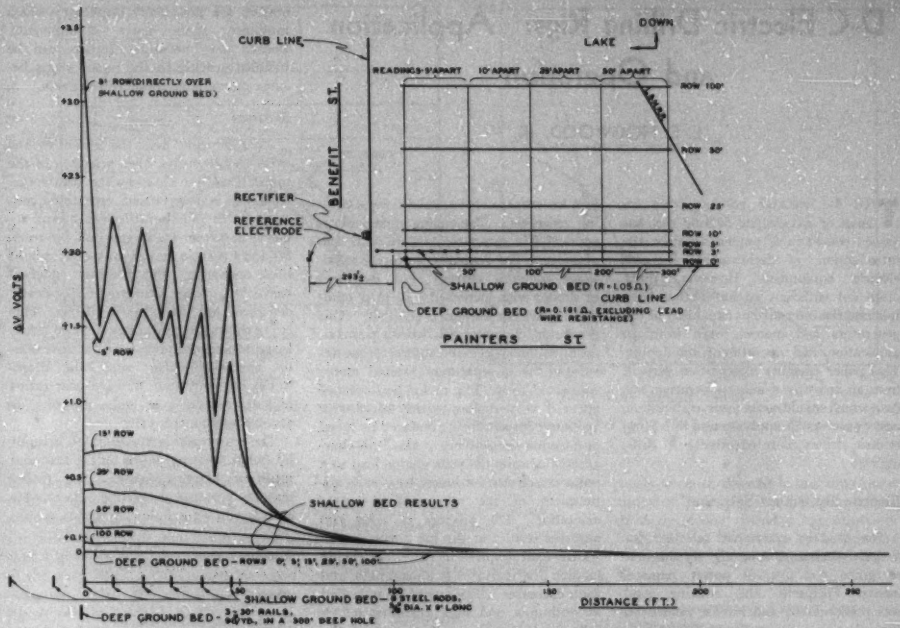


Fig. 9. Voltage gradient comparison of deep and shallow ground beds at Painters and Benefit Streets

Rectifier settings for deep ground bed: I = 9.75 amp; E = 3.0 volts (d-c)
Rectifier settings for shallow ground bed: I = 10.1 amp; E = 10.7 volts (d-c)

Table II. Resistance to Earth of Various Deep Rail Ground Beds

Location and Date Installed	Description*	Resistance to Earth, Ohms
Valence Substation no. 1	(A)	0.047
Valence Substation no. 2 (4–15–54)	(A)	0.049
General Pershing near Rendon (12–18–53)	(B)	0.217†
Roman near Tulane (12–21–52)	(B)	0.251†
Canal Boulevard near Robert E. Lee (12–22–52)	(B)	0.233†
Bienville Plaza no. 1	(C)	0.334†
Bienville Plaza no. 2	(C)	0.364†
Bienville Plaza no. 3	(C)	0.251†
Combination of nos. 1, 2, and 3 (9–10–54)		0.110†
Washington Square (5–24–54)	(D)	0.055

* A—360 ft of 100-pound-per-yard T-rail, top flush with ground, top 60 ft coated.
B—Two 30-ft lengths of 100-pound-per-yard T-rail, bottom 250 ft deep.
C—Four 30-ft lengths of 100-pound-per-yard T-rail, bottom 390 ft deep.
D—360 ft of 100-pound-per-yard T-rail, top flush with ground, all bare.

† Includes resistance of no. 6 (or no. 8) wire from rectifier to rails (R/1,000 ft no. 6 = 0.394 ohm; no. 8 = 0.627 ohm).

earth would be 0.17 ohm. Two rails side by side, buried horizontally 9 ft deep, would be 0.5 ohm, so it can be seen that in the New Orleans area extremely low resistance to earth can be secured when the rails are located deep.

Fig. 9 shows, in one illustration, a number of changes of earth potential caused by 10-amp discharge from a shallow bed and later from a deep rail ground bed, the ground beds not being energized simultaneously. Change of potential difference readings between a $CuSO_4$ electrode at various intervals and a remote $CuSO_4$ electrode were measured in a line directly over the shallow bed, and in rows 2, 7, 22, 47, and 97 ft away from and paralleling the shallow ground bed. Another set of readings was made directly over a deep rail ground bed and in rows 5, 10, 25, 50, and 100 ft away from the deep rail ground bed. The results of these tests were plotted as shown, and demonstrate that even for 10-amp discharge in 500-ohm-cm soil, appreciably steep earth gradients are encountered for the shallow bed, whereas for the deep bed, the curve is flat.

Conclusions

The advantages of deep ground beds may be briefly summarized as follows:

1. Full rectifier capacity is available for protection of structure to be cathodically protected. No current need be drained from foreign lines because of direct pickup effect. This means elimination of distributed drainage system for foreign lines, with consequent savings.

2. Lower power costs result because of drastically reduced ground bed to earth resistance.

3. Smaller space is required for a deep ground bed.

4. Deep ground bed can be installed at any convenient location. No right-of-way problems are involved.

5. Cost of cathodic protection coordination tests is practically eliminated since current pickup on foreign lines is negligible.

6. No time is lost in surveys to find suitable locations for conventional shallow ground beds.

D-C Electric Drilling Rigs: Application and Operation

E. E. HOGWOOD, JR.
ASSOCIATE MEMBER AIEE

THE d-c variable voltage drive for diesel or gas-electric drilling rigs has gained considerable popularity since the introduction of lightweight low-cost electric equipment. However, quite often not sufficient advantage has been taken of the intermittent capability of the generators and motors, both from an application and operational standpoint. This paper presents information derived from an analysis of actual operating test data which should assist users, rig builders, and operators in applying and operating electric drives more effectively on drilling rigs.

Electric Equipment Selection

The electric equipment selection for a diesel-electric drilling rig should first recognize the over-all power requirements. Formerly, the hoisting load was predominant and its size influenced the selection. On modern drilling rigs, the mud pumps are the largest single load; thus, their horsepower requirements are a deciding factor in the choice and number of engine-generator sets for the power plant. Two, three, or four sets are usually selected depending on engine size, degree of reliability desired, and the load requirements. Once the engines are selected, the generators, motors, and control should be chosen. This paper starts with the generators since they are the first link in any electric drive system or transmission.

Generators

The first requirement in the selection of the d-c generator is that it have characteristics which will suitably match the engine. An engine is essentially a constant-torque machine from which variable hp (horsepower) can be obtained by speed variation. Its rated horsepower is obtained at maximum rated speed; therefore, it is important that the generator match the maximum continuous hp and speed of the engine. Second, the generator output volts and ampere rating or some multiple of either must also match those of the motors being considered.

Fig. 1 shows the desirable volt-ampere and kw-ampere characteristic for a drilling generator. The engine horsepower is expressed in generator output volts and amperes. The generator voltage is maximum at zero current and decreases or droops with increased load to a minimum when the current is greatest. This generator characteristic fixes the maximum no-load speed and stalled torque developed by a separately excited motor connected to it. The motor performance curve is very similar to this curve with volts and amperes being replaced by speed and torque respectively. Also, this characteristic limits the peak engine load to a value which can be preset by simple adjustment of the maximum generator excitation. The product of volts and amperes results in the kw curve shown. As the load amperes increase, the engine loading first increases to a maximum, and then decreases. The engine is unloaded at both low and high currents and is fully loaded on a small portion of the curve. Therefore, the generator should be selected so that its peak kw exceeds slightly the continuous engine output. Five- to ten-per-cent overload on the engine at peak kw is permissible and desirable when required. Overloading allows the system to operate at or near peak power over a wide load range.

The electrical rating of the generator is established on the basis of a continuous ampere load and a permissible temperature rise. It is usually not possible to take advantage of the full intermittent capability of the generator even on cyclic loads because it is desirable to limit the maximum loading on the engine. The generator kilowatt rating converted to horsepower should match the continuous output of the engine.

During recent years available generator frames and ratings have led to the application of one or two types of generators to a wide range of engine sizes. This procedure has many advantages for both the user and the electric equipment manufacturers. First, it offers low-cost low-weight electric equipment on a short delivery basis. Second, it is possible to adjust the generator peak output so that a number of different oil field engines can be matched. Third, the motors and generators are practically duplicates which minimize spare parts requirements. Fourth, any necessary repairs can be handled quickly in the repair shops because of available standardized parts.

Motors

Consideration must be given to the drive requirements when selection of the motor is made. Consider the drawworks. This load is intermittent, essentially constant hp in the operating range, and requires a certain maximum pull for stuck bit and a certain maximum speed for light hook operation. Fig. 2 shows a typical curve for a large drawworks specifically designed for an electric drive. This curve shows the performance of two 1,000-hp motors receiving power from three 800-hp engine-generator sets. The drawworks has only four forward gear ratios and no reverse gear since reversing is accomplished electrically.

The drawworks is usually rated in input hp (while hoisting) based on the load and resultant hoisting speed. This rating may be expressed as hoisting power and is calculated by the following equation.

$$\text{Hoisting Power} = \frac{P_L \times N_L}{33{,}000 \times E_{DW}}$$

where

P_L = fast line pull in pounds
N_L = fast line speed in fpm (feet per minute)
E_{DW} = efficiency of drawworks

Since the load requires this hoisting power for a variable length of time after which there is a semifixed interval of time of no-load for "breaking-out" the stand and setting it aside, the drawworks motor may be applied with a continuous hp rating considerably below the drawworks hp rating.

The amount of power that the motor can transmit to the drawworks on an intermittent basis depends on: its designed capabilities from a maximum torque and maximum speed standpoint; and the amount of power available from the engine-generator sets. The motor is designed to carry a continuous current under load conditions without exceeding a maximum safe temperature. Applied to the drawworks, it does not carry current for the entire hoisting cycle; there-

Paper 59-1103, recommended by the AIEE Petroleum Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Petroleum Industry Conference, Long Beach, Calif., August 23-26, 1959. Manuscript submitted November 10, 1958; made available for printing July 3, 1959.

E. E. Hogwood, Jr. is with Westinghouse Electric Corporation, East Pittsburgh, Pa.

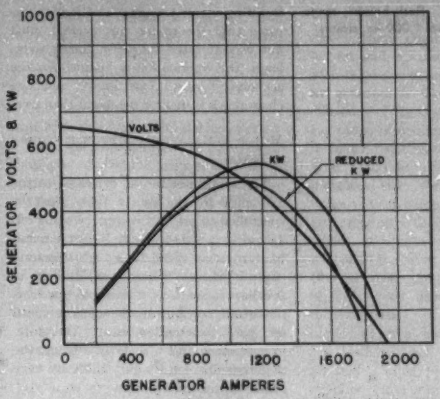


Fig. 1. Performance curves of one 550-kw d-c generator for use on drilling rigs

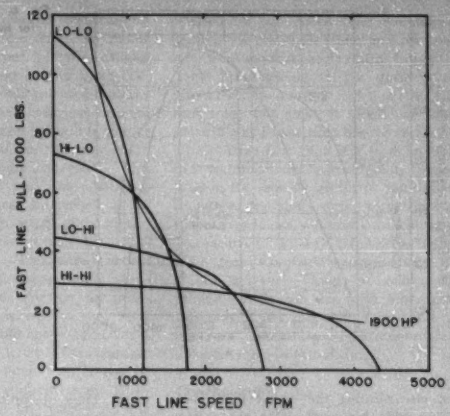


Fig. 2. Performance of two 1,000-hp motors on a large drawworks

fore, it can carry considerably more than rated current during the hoisting portion without causing any thermal damage to the insulation. The maximum current that a given motor can carry, regardless of the duty cycle, is determined by commutation. Also, the maximum voltage which can be applied to the motor is determined by the point where either flashing-over or serious overspeeding occurs. Both of which are undesirable.

A motor then can receive power from one or several engine-generator sets and deliver more output to the drawworks on a short time basis, as long as the safe maximum current and voltage ratings are not exceeded and the effective current is held to rated amperes or less.

The effective current which establishes the amount of heating that the motor can withstand is defined as the rms of the instantaneous currents for a given duty cycle. For instance, Fig. 3 shows the current trace of two drawworks motors connected in series for two successive hoists. Hoist no. 1 is in Low-High gear hoisting 5-inch drill pipe from 5,590 feet. Hoist no 2 is in High-High gear hoisting essentially the same load from 5,500 feet. The rms amperes for each cycle is determined by using a mean value of current I_m and assuming this current is required during the hoisting portion of the cycle. This is an approximation, but is close enough for application purposes.

$$\text{rms } I = \sqrt{\frac{I_m^2 \times t_h}{t_h + t_r}}$$

where

I_m = mean amperes during hoisting portion of cycle
rms I = effective current in amperes
t_h = hoisting time in seconds
t_r = rest time in seconds

For hoisting cycle no. 2 of Fig. 4:

$$\text{rms } I = \sqrt{\frac{(1{,}300)^2 \times 23.3}{23.3 + 33.3}}$$

$$= \sqrt{692{,}000}$$

rms I = 832 amperes

This calculation is based on force-ventilated motors. If the motors were self-ventilated, the rest time would be reduced by a factor of 4 because the heat cannot dissipate as rapidly.

Each motor on which this data was taken is rated 960 amperes continuous; therefore, the peak current of 1,470 amperes and the "running" current of approximately 1,350 amperes for this hoist is satisfactory.

On the rms-ampere basis, higher currents could have been taken during this hoist, because the rms is less than rated amperes; however, it will be shown later that the driller obtained approximately maximum hp transfer to the drawworks by using a running current of 1,350 amperes.

Motor Application Factor for Drawworks

An analysis of test data taken over a period of many years indicates that approximately 85–90% of the hoisting cycles are in the 20–60% "on" classification. In other words, the motor is delivering power to the drawworks from 20% to 60% of the total time per stand and is resting from 80% to 40% of the total time. Of course, the time "on" per stand varies with the amount of power on the rig, the load, and the gear ratio; however, a conservative cycle is estimated to be 50% on and 50% off. The hoisting motor gains additional cooling after completing a round trip because it is usually at rest with full ventilation

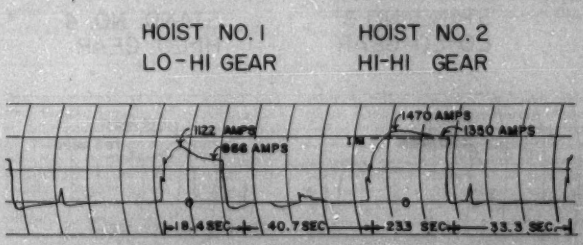


Fig. 3. Amperes versus time for two successive hoists by two 1,000-hp motors

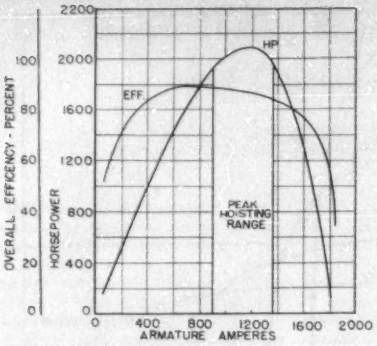


Fig. 4. Peak hoisting range for two 1,000-hp motors

or else rotating the bit at light load. This adds conservatism to the factors used in determining the actual hp rating of the motor for the drive.

Based on the preceding, it is standard practice to apply drawworks motors so that they develop approximately 140–145 % of their continuous rating on a 50-50 hoisting cycle. This is conservative enough to handle normal hoisting without allowing the rms current to exceed rated current. An application factor of 1.4–1.45 for the drawworks motor means that one 1,000-hp continuous-rated electric motor can drive a drawworks with a rating of approximately 1,400–1,450 hp if the chain ratios are properly selected and sufficient power is provided.

Motor Application Factor for Mud Pumps

The mud pumps are designed for an input hp based on a maximum speed in strokes per minute and a maximum force in pounds. Built-in protection in the form of a pressure relief valve or shear pin is usually provided by the pump builder to protect the pump against overstressing the mechanical parts when high loads exist. Since this pressure relief is usually set at approximately 25-30% above rated pressure, the electric drive system should have built-in torque limit to prevent unnecessary operation of the pump protective device under these conditions. This is accomplished by the generator characteristics as previously discussed.

The running load of the mud pump is sufficiently continuous to warrant an application factor of 1 for the mud pump motor. This means that a 1,000 hp motor should be applied on a pump requiring 1,000 hp rated input. For reasonably long periods the pump will be operated at full load and speed. Strictly speaking, however, the pump is not running on a continuous basis for longer than approximately 2 hours; therefore, by applying an electric motor at the same rating as the pump input rating, a conservative margin exists to provide extra capacity when an emergency demanding an overload arises.

It is usually more economical to apply one motor than two on any large drilling rig drive. In some instances a motor has been applied to a pump whose rating is higher than the motor. If the difference in rating is not too great and it is recognized that the motor rms current must not exceed rated amperes during overloads, this application is justified for the economical gain. One motor and one chain drive is more economical than two.

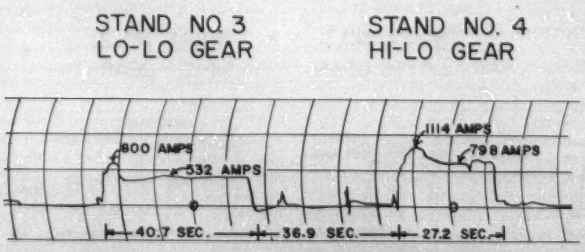


Fig. 5. Amperes versus time for stands 3 and 4 from 10,000 feet

Motor Application Factor for Rotary Table

The maximum torque delivered to the drill pipe by the rotary table must be controlled to prevent it from twisting off. Likewise, it is desirable to limit the maximum rotating speed to provide adequate control when "fishing" or drilling in peculiar formations. Variable speed is important to provide the correct speed for best penetration rate. Therefore, from a speed and torque standpoint, the requirements of the rotary table are very similar to the mud pump.

The running load of the table, like the mud pump, is sufficiently continuous to warrant an application factor of 1. It may be driven directly by a separate motor or from the drawworks. If separately driven, the maximum torque supplied by the motor and the chain ratio from the motor to the table may be selected so that the torsional limit of the drill pipe will not be exceeded. If the table is driven from the drawworks, some means of torque adjustment on the drawworks motors must be provided to provide this protection because the rotary table load is relatively low in comparison to the hoisting load. This can be accomplished automatically when the driller operates the "Drill-Hoist" selector switch by reducing the number of generators available to the drawworks motor when drilling or by inserting resistance in the generator field circuit to limit the excitation.

Control

The selection of the control system depends on several considerations. The first is the degree of power flexibility required on the rig. The control can be arranged so that the generators may be switched to any motor on the rig or to only one motor. Of course, these are extreme examples of flexibility; however, some operators have deemed it justifiable to provide 100% flexibility. For most rig applications some intermediate degree of flexibility is usually satisfactory.

Certainly, all or almost all the generators on the rig should be connectable to the drawworks motor(s) and perhaps to one large pump. This will provide full power should any one of the engine-generator sets fail. Of course, the degree selected depends on the user's desire and

the amount of insurance required. For instance, offshore, the additional expense of having each motor connectable to all generators is usually justified as is extra generating capacity in the power plant on the basis that down-time is much more expensive than the additional investment.

A second consideration is the over-all simplicity of the control system. Most operating and maintenance personnel are not sufficiently trained as yet to handle the more complex schemes available to the electrical design engineer today. For that reason the control system should be as simple as possible and yet provide the necessary operational and protective features for the equipment and personnel.

Simplicity means the desirability of having the control and machine characteristics as much a part of the inherent system as possible without additional rotating machines, relays, regulators, air systems, slide valves, air actuators, and systems requiring critical adjustment. With characteristics built in the main generators or into static, easily-adjusted systems, the operating personnel have a minimum of concern about how to find and fix a trouble because troubles are minimized. With the saturable reactor, magnetic amplifier, silicon rectifiers, and other "no-moving-parts" devices, simple, easily understood control systems can and are being built today to minimize down-time due to component failures in the control system. The machine tool industry has found these systems invaluable from a maintenance standpoint and there is no reason why such systems cannot be applied on drilling rigs to provide improved control systems.

Improvement of Operation

Since the motors and generators serve as an electric torque converter between the engine and driven equipment, it is well to analyze their performance for methods of improved operation. As noted earlier the d-c generator automatically limits the peak loading on the engine by limiting its own maximum output.

Fig. 4 shows the over-all electrical efficiency from generator input to motor output and hp output for two 1,000-hp motors receiving power from three 550-kw generators. It will be noted that the efficiency of the system is highest through the range of maximum hp, and is similar in many respects to the efficiency curve of a hydraulic torque converter. This indicates that the optimum operating range is near the peak kw output of the generator or peak hp output of the motor, where maximum power will be transmitted to the drive. This range is shown between 900–1,350 amperes and is intended to be used only on an intermittent basis. It is not to say that higher amperes should not be used, for if the load requires 1,700–1,800 amperes for high torque pulling, then it can be supplied on a momentary basis. However, for maximum hp transfer to the hoist, the running current should be between 900 and 1,350 amperes for the motors in this example.

This is particularly important on the drawworks during hoist because a significant saving in time can be made if the driller changes to a higher speed gear ratio at the right time. Not only will the equipment be operated near peak efficiency but it will also transmit peak hp from the engines which will result in minimum hoisting times. Such operation is best for the electric equipment, because it will not be overloaded on an rms basis, but will allow full utilization of its intermittent capacity at peak performance on intermittent loads.

A suggested method of operation for the driller is to have the drawworks ammeter marked in green for the peak hp range of currents, red above peak hp marked "Shift Down" and yellow below peak hp marked "Shift Up." In this manner, the driller will not have to make a test shift when he thinks the hoisting speed can be improved. It can definitely be improved with one shift if the ammeter markings are made taking into consideration the motor performance and the gear ratios of the drawworks.

On some recent tests in the Gulf of Mexico, it was noted that the driller hoisted the first three stands in Low-Low gear from approximately 10,000 feet with 2,000 hp available to the drawworks. From the data shown in Fig. 5 of stands 3 and 4 for this hoist, it is evident that stand 3 could have been hoisted in High-Low or even Low-High gear from an ampere and maximum hp standpoint. Also, later in the same "round-trip" the driller made two test shifts trying to get into a higher speed ratio, whereby some lost time was incurred. Finally, the shift to High-High gear was made approximately 10 stands too early, losing an additional minute. Although the total time lost was only a few per cent of the total, a number of round trips could account for a considerable length of time.

Conclusions

In summary, the application of electric equipment to drilling rig drives can be made more economical if closer consideration of the engine, load, and electrical equipment is made. The generator should provide as close a match as possible with the engine to transmit maximum hp at highest efficiency. The motors should be selected on the basis of the load so that in cases such as the drawworks, the intermittent capability of the motor can be used. On the mud pumps and the rotary table, the maximum torque and continuous hp requirements must be considered. Selection of the control system should be made on the basis of simplicity with a minimum of parts which provides the desired degree of flexibility and protective features.

Consideration should also be given to the method of operation used by the driller. For peak hp transfer and optimum operating efficiency, the driller should shift gears on the drawworks at the time where maximum benefit can be gained. This point will be considered when the application is made if the rms capacity of the motor is utilized. By recognizing these points and taking advantage of them, the builder, user, and operator of modern electric rigs can all benefit.

Discussion

W. E. Withers (Sun Oil Company, Beaumont, Tex.): I want to compliment Mr. Hogwood on his investigation and analysis of the various loads on d-c electric drilling rigs. We are in complete agreement that the user has not realized the full advantage of the electric drive as far as the intermittent capabilities are concerned. Too many users do not understand the theory, operation, and flexibility of electric drives; thus, when installing equipment to drive a drawworks, the installed capacity of the motors are equal to or closely approaches the rated capacity of the drawworks. This way of thinking has stemmed from the use of internal combustion engines for these drives. The manufacturer as well as personnel within the user's organization is responsible for recommending the proper sized equipment.

The author points out that the generator kilowatt rating converted to horsepower should match the continuous output of the engine. This is the ideal condition but somewhat impractical economically. When using the Ward-Leonard system of control with five separate loads, it would be necessary to have five engine generator sets. All of these could be the same size but would have to be reasonably large to accommodate the mud pump loads. I believe an installation with two generators on each engine can be used very successfully if proper con-

sideration is given to the division of load. This arrangement will reduce the number of engines and consequently the area required for this equipment.

The point made by the author concerning standardized parts, fast delivery, and quick repairs for the equipment is very necessary in the petroleum industry inasmuch as time is very costly.

Mr. Hogwood uses the term effective current (rms amperes) in determining the load on the drawworks motors. This method of calculating the effective current gives some meaning to the term "intermittent duty" which so many manufacturers use. Inasmuch as the loading of a motor is primarily governed by the ability of the motor to dissipate the heat generated within it, the quantity of forced ventilation is also a factor that must be considered. However, with ventilation as supplied by the manufacturer, Mr. Hogwood's calculations of the rms current to determine maximum loads for cyclic loading is quite satisfactory.

The rotary table load is the smallest drilling load encountered, but a very important one. The size of this load is influenced by the drilling program, subsurface formation, and hole condition. A close approximation for the rotary table load is one horsepower per revolution of the rotary table; however, with a high rate of penetration and a large hole (18 inches) the table motor can be stalled. Further investigation of this load is needed by both the manufacturer and the user.

The author's discussion of control simplicity is of utmost importance. Any and all control circuitry should be as simple and straight forward as possible with a minimum of equipment requiring critical adjustment.

A. H. Candee (The White Motor Company, Pittsburgh, Pa.): Mr. Hogwood has presented a paper which is worthy of considerable thought. His analyses and conclusions are based upon a background of experience with all types of electric drives, backed by actual field tests. It is upon such foundations that progressive evolution rests, and it is to be deplored that many rather recent applications of electric drives have been made hastily under the pressure of competition, and by their shortcomings and failures have done much to cast undeserved discredit on electric drives as a whole.

There is little in Mr. Hogwood's paper which warrants adverse criticism, but there are a few points which, perhaps, may be expanded or emphasized.

The paper indicates the need for control simplicity and practicality, but it is thought that this deserves added emphasis. Mention is made of the flexibility of electric drives which permit any generator or combination of generators to be used to feed any motor. While this is true, the designers of electric drives often go too far in this respect, for each time a different combination is added, control apparatus and electric contact-making devices must be added, yet the failure of any control circuit device to make an electrical contact when it is supposed to do so may leave the driller and his crew completely lost if the electric control circuits become too complicated. For this reason, the electric circuits must be kept severely simple. While it is true that protection must be arranged to take care against the loss of one or more engines, the degree of protection against the loss of a generator need not be nearly as broad, for the modern electric generator or motor is far more reliable than any gas or diesel reciprocating engine now in use for the drilling of oil wells. In addition to keeping the electric circuitry to its simplest terms, the designer must also avoid any and all apparatus which will not stand up without question in the extremely rough service encountered under normal drilling conditions. Flimsy contact arrangements, fine-wire devices, fragile mechanical structures, and devices which need careful adjustment are to be shunned.

In Fig. 1, Mr. Hogwood shows a typical volt-ampere curve for an electric drive, with the corresponding kilowatt (or horsepower) curve. Fig. 2 then shows how an electric drive requires but four gear ratios in the drawworks. To some of us this points out the need for further work toward utilization of the available engine power over a much wider range of speed so that the number of speed ratios which must be built into the drawworks may be still further reduced, with resultant reductions in first costs and maintenance expenses and with an improved reliability factor.

The paper, as well as Figs. 1, 3, 4, and 5, is specific as to ampere values. It should, of course, be fully understood that these values apply only to the Westinghouse Type *371* generators and motors and not to motors of other manufacture. If competitive motors are compared on an ampere basis, the results may be very misleading. This is due, of course, to the fact that without a fixed voltage or frequency to shape the design, the designer may choose the voltage and ampere capacity best suited to his manufacturing facilities. Thus, motors of equal capacity and stamina built by different builders may differ 20 to 30% in ampere ratings.

There is much work which may yet be done to improve diesel-electric drive systems. It is papers such as this which will assist materially in pointing the way toward increased reliability and efficiency.

G. W. Webb (General Electric Company, Dallas, Tex.): Mr. Hogwood's paper presents a sound approach to the subject of application of d-c drives to drilling rigs. The basic principles covered have been widely accepted in the industry, particularly since the advent in 1955, of low-cost, lightweight, short-delivery modified traction machines.

Such machines by their nature emphasize the principles stated by the author of:

1. Standardization on fewer ratings
2. Engine overload protection
3. Motor torque and speed limits
4. Simple controls
5. Utilization of intermittent capabilities.

By way of comment on these principles, some 15 manufacturers offer about 30 different horsepower ratings of drawworks from 100 hp to 2,500 hp, and about 40 different horsepower ratings of mud pumps from 100 to 1,700 hp. The inherent inefficiencies of such an equipment supply situation are obvious when considered on an industry basis. All these models must be redesigned periodically to keep up with competition and advancing technology. As a result, few units of any one make and model can be sold to the limited market available. The economic advantages of mass production are seldom obtained and much of the equipment cost goes to cover development and engineering.

The electrical manufacturers so far have each offered only two or three ratings of motors to drive the wide range of drilling machinery and two or three ratings of generators to be driven by the even wider range of engine horsepower ratings and speeds available. The economic benefits of this approach must be sold to the drilling contractors and oil companies who specify equipment. It could result in future savings to all parties involved.

The author points out that peak engine loads are limited by the differential generator characteristic, but recommends that the generator kw rating should closely match the continuous output of the engine. It has been demonstrated that a 3-field generator, including a self-excited shunt field, can be applied to engines of much lower horsepower capability than the generator rating and still prevent engine overloading or stalling. By flattening the top of the hp-versus-motor-rpm curve, this device enables the electric transmission to transmit full engine horsepower to the load over a wider range of load speeds.

The author mentioned the inherent torque and speed limits on separately excited motors by virtue of the generator characteristic. He indicated that the torque limit could be changed by limiting generator excitation or reducing the number of generators available to a motor. One other device which can be used to limit torque applied to a pump or rotary is to weaken the fixed motor field. This trades torque for speed without reducing horsepower, but the drive ratio must be selected accordingly.

We concur with the author's remarks regarding control simplicity. Basic drive performance should be inherent in the rotating machine characteristics. We feel that complicated voltage and current regulating systems are best avoided by using a differential generator and that this electrical characteristic is best translated into controlled speed and torque by the separately excited shunt motor. We would like to inquire about the author's experience regarding the long term reliability of rectifiers in view of the a-c voltage fluctuation sometimes encountered on drilling rig light plants.

Regarding application factors for drilling motors, we note that the author recommends a factor of 1.4–1.45 for the drawworks, whereas his company's advertisements specify a 1.6 factor. We have found the 1.6 factor to be conservative for this application. Referring to the section of the paper on "Motor Application Factor for Mud Pumps," I would emphasize the last paragraph which implies that economic consideration may dictate an application factor exceeding unity. Where experience is available to predict accurately the pumping load cycle, a factor in the order of magnitude of 1.1 or 1.2 may apply. This usually is justified on the basis of possibly increased motor maintenance versus increased initial investment.

In closing, I would like to comment on the general aspects of the application of electric transmission to drilling machinery. The

author describes a drawworks with four forward mechanical ratios and no reverse gear as being "specifically designed for electric drive." This is true when compared to conventional, straight mechanical drawworks which normally provides 6 or 8 forward ratios plus reverse gear and friction clutches. However, by more fully utilizing the torque conversion characteristics of the d-c transmission system, the drawworks can be further simplified. It is quite feasible to design a drawworks specifically for electric drive with positive clutches and two mechanical ratios to the drum shaft without sacrificing essential hoisting performance. Such a drawworks could offer the user savings in weight, size and initial cost which would provide the economic and performance benefits of electric rigs at no greater first cost than mechanically driven rigs.

E. E. Hogwood, Jr. The author wishes to thank Mr. Withers, Mr. Webb, and Mr. Candee for their discussions and comments.

This paper principally outlines application procedures that are presently used in other industries where electric power is more widely used. As the drilling industry becomes more familiar with diesel-electric power, it is expected that the users and suppliers will learn to use these principles to utilize fully the electric equipment. The result will be a reduction in first cost and operating expenses of drilling oil wells using electric drives.

First, I want to answer the specific questions that were asked and then, briefly, comment on some of the points common to more than one of the discussions.

1. This paper was presented as a conference paper at the 1958 Petroleum Industry Conference and since that writing, a considerable amount of development work and study of field test data have been made. On the basis of these data, we have initiated and completed development work on a new static excitation and control system for oil well drilling. It is still agreed that the use of inherent machine characteristics is a good, simple method of obtaining the required drive performance. We have used the 3-field generator or exciter and the separately excited shunt motor for more than 20 years to obtain this performance for drilling; however, the static excitation system has been developed in an effort to minimize cost further, to increase the output of a given generator and motor frame, to simplify the system, and also to maintain the over-all system reliability. This system utilizes static components such as magnetic amplifiers, saturable reactors, and silicon diodes which need no attention once the system is set. This system is considered to be a definite improvement over the already proved system which relies on inherent machine characteristics.

2. As for the long term reliability of rectifiers considering fluctuation of a-c power on the rig, we have not experienced any difficulty attributable to this cause. However, selenium rectifier applications should be made rather conservatively to avoid excessive aging. Some of the factors affecting aging that should be considered are high ambient temperatures, corrosive atmospheres, cyclic overloads, and long cable runs to the loads. Where silicon rectifiers and diodes have been used, no aging difficulties have been encountered and they are proving to be quite reliable.

It is apparent also that all a-c auxiliary power sets used on drilling rigs should be equipped with reasonably good voltage regulators to operate motors, controls, lamps and other loads within rated tolerances.

3. The 1.6 application factor for the drawworks drive was published by us after this paper was originally written including the 1.4 to 1.45 factor. The application factor depends on the length of the hoisting cycle to rest cycle ratio which is based on the rms current not exceeding the rated current of the motor. It should be noted that a shorter hoisting cycle than that mentioned in the paper is required for the 1.6 factor than for the 1.45 factor. It has been found that the 1.6 ratio is conservative for this application. It is suggested that the Petroleum Industry Committee consider establishing standards for the presently used "intermittent rating" based on rms current and an acceptable hoisting cycle. The basic need is for a standard hoisting cycle which could be used to determine the drawworks application factor for various manufacturers equipment.

I concur with Mr. Withers that applications of two generators per engine offer an economic means of dividing engine horsepower; however, the maximum rating of either generator should still not exceed the horsepower capabilities of the engine. Some engine manufacturers prefer each generator to be capable of absorbing the full output of the engine, even though both never could do so simultaneously. This allows utilization of the full engine output through one generator when required.

Mr. Webb has stated that a 3-field generator, including a self-excited field, can be applied to engines of much lower horsepower capability than the generator rating. This is basically true, depending on the ratio of separately excited field strength to self-excited field strength in the generator design; however, it is felt that a wider range of constant horsepower can be accomplished more economically in the control system than in oversizing the generator.

One more point in closing. Mr. Webb and Mr. Candee both discussed the point concerning a 4-speed drawworks specifically designed for electric drilling. To their comments, I would add that I fully agree and that I hope we may soon be requested to study and build an electric drawworks as Mr. Webb described with some major rig builder. By reducing the mechanical equipment required and fully utilizing the electric drive, certainly the first cost of the electric rig will be approximately equal to the mechanical rig. With the first cost equal, the user will be more desirous of having the electric drive with its additional advantages.

Three-Phase Induction Heating Coils

N. V. ROSS
AFFILIATE AIEE

IN THE 60-cycle field of induction heating a 3-phase system is desirable. The advantages of 3-phase power is fairly obvious. In most cases a large amount of power is required and a balanced 3-phase load is very practical. Hence, 3-phase induction heating coils have been used to a large extent in 60-cycle heating of aluminum and brass. Since the size of stock to be heated is large in cross section, which is the primary reason for 60 cycle, the amount of energy required to heat the material because of its large mass is, in many cases, quite high. In the past, large billets, 4 inches to 30 inches diameter of aluminum, as well as brass, 4 inches to 10 inches in diameter, have been heated primarily for extrusion. Temperature uniformity is mandatory for high-quality results. In the case of special brass alloys, the temperature must be maintained at ±10 F (degrees Fahrenheit) throughout the billet. Induction heating is primarily a very rapid heating process taking seconds or minutes to heat, while other methods of heating require hours. Because of this rapid heating, temperature patterns follow closely magnetic flux patterns; also due to the rapid heating, sufficient time is often lacking for equalizing.

As is well known, the magnetic flux pattern described by Fig. 1 is fairly uniform throughout most of its length for long coils. The field intensity H diminishes at the ends to one half the value at the center of a closely spaced solenoidal coil. If a long cylindrical billet is placed inside the coil, the temperature or heat pattern will follow the field intensity pattern, that is, if the billet is longer than the

Paper 59-256, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted November 26, 1958; made available for printing January 28, 1959.

N. V. Ross is with the Magnethermic Corporation, Youngstown, Ohio.

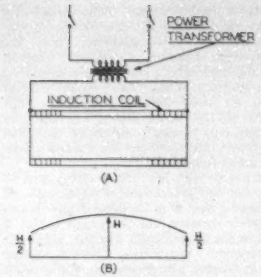


Fig. 1. A—Single phase coil. B—Corresponding field intensity

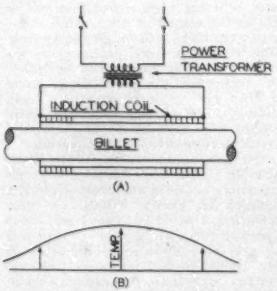


Fig. 2. A—Single phase coil with long billet. B—Corresponding temperature pattern

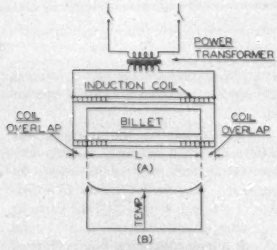


Fig. 3. A—Single phase coil with short billet. B—Corresponding temperature pattern

coil; see Fig. 2. If the billet is shorter than the coil, the ends will tend to overheat as shown in Fig. 3, because of the concentration of flux lines at the ends. The excess coil length is known as coil overlap. The coil overlap depends upon several factors; the material of the coil (in most cases copper), the coil space factor, the material of which the billet is composed, the frequency applied to the induction coil, the inside diameter of the coil, and the diameter of the billet. At present the amount of coil overlap using 60 cycles per second is determined primarily by experiment and varies from 3/4 inch to 2 inches for aluminum billets, and $1^1/_2$ inches to 3 inches for brass. It is fairly easy to obtain a uniformly heated billet with a coil as illustrated in Fig. 3 with single-phase power. The ends are purposely overheated during the heat stage to compensate for radiation losses from the edges as the billet travels from the heater to the press.

In the case of 3-phase coils a phase dip occurs in the flux pattern as shown by Fig. 4. Since the heat pattern is a function of the field intensity squared, the phase dip is quite pronounced. The first improvement was realized by reversing the phase of the center coil section to obtain 60 electrical degrees displacement between phases instead of 120 degrees. This is shown in Fig. 5. The dip in field intensity H at the phase points became less and the corresponding temperature pattern was improved. The results were still not good enough and a better solution was sought. The answer was one of phase overlapping as illustrated by Fig. 6. At the phase points, a small number of turns were simultaneously energized by adjacent phases. The number of turns or phase region overlapped allowed one to obtain nearly uniform flux pattern throughout a 3-phase coil approximating the single-phase coil. The corresponding temperature pattern was improved to such a degree that a billet can be heated within ±10 F throughout its entire length. The amount of phase overlap required depends upon several factors as in the case of coil overlap. It varies with coil diameter, billet diameter, frequency, billet, and coil conductor material. The phase overlap can be calculated by the following equation:

$$\phi_{OL} = 1/4[(b+d_1)-(a-d_2)] \qquad (1)$$

where

ϕ_{OL} = phase overlap
b = coil inside diameter
a = billet outside diameter
d_1 = penetration depth of coil conductor
d_2 = penetration depth of the billet

If b and a are measured in inches, then d_1 and d_2 for nonmagnetic materials are calculated by:

$$d_1 = 3{,}160\sqrt{\frac{\rho_1}{sf}} \text{ and } d_2 = 3{,}160\sqrt{\frac{\rho_2}{f}} \qquad (2)$$

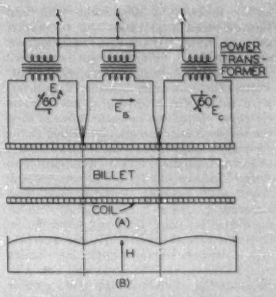


Fig. 4. A—Transformer and coil connections 3-phase nonreversed center section. B—Corresponding field intensity for coil as in (A)

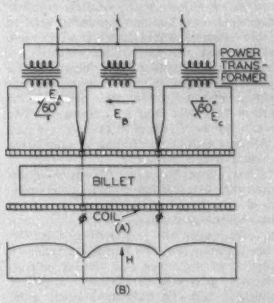


Fig. 5. A—Transformer and coil connections 3-phase reversed center section. B—Corresponding field intensity for coil as in (A)

where

ρ_1 = resistivity of coil conductor in ohm inch
ρ_2 = resistivity of billet material in ohm inch
s = coil space factor
f = frequency cycles per second

Equation 1 holds for 3-phase coils only. Since an integral number of turns can be contained within an overlapped phase region, the nearest whole number of turn widths is usually used for the phase overlap calculated by equation 1.

Example: Calculate the phase overlap of an aluminum billet of 5-inch diameter and length 24 inches, coil dimensions; 6-inch inside diameter and 26-inch length. The copper turn width is 3/8 inch with 1/32-inch turn insulation. Coil space factor = 0.92. Mean temperature of aluminum to be 850 F, with 60-cycle 3-phase power supplied to the coil.

Solution: First calculate the penetration depths for the coil copper and the aluminum load. Use average resistivity

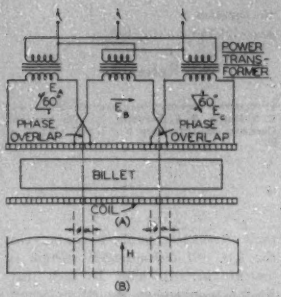


Fig. 6. A—Transformer and coil connections 3-phase reversed center section with phase overlap sections. B—Corresponding field intensity for coil as in (A)

values over the temperature range. The average resistivity of copper is 0.8×10^{-6} ohm-inch.

Hence:

$$d_1=3{,}160\sqrt{\frac{0.8\times10^{-6}}{0.92\times60}}=0.380 \text{ inch}$$

and

$$d_2=3{,}160\sqrt{\frac{2.24\times10^{-6}}{60}}=0.611 \text{ inch}$$

$$\phi_{OL}=1/4[(6+0.38)-(5-0.611)]$$

$$\phi_{OL}=1/4[(6.38-4.39)]=0.498 \text{ inch}$$

The total turn width is $0.375+0.0314=0.406$ inch. Hence, two turn widths would give 0.812 inch. In this case either one or two turns could be chosen but for physical space reasons, two turns would be used.

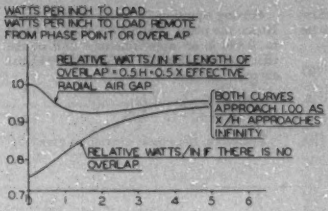


Fig. 7 (left). Ratio x/h of distance from phase point or center of overlap to effective radial air gap

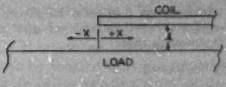


Fig. 8 (above). Diagram for calculating field intensity near end of long cylindrical coil

Conclusions

One may conclude from the preceding that it is possible to obtain a uniform temperature pattern over the length of an inductively heated billet using a 3-phase coil. A slight dip or rise in temperature, approximately ±10 F, still exists at the phase points due to the necessity of using an integral number of turns approximately equal to the calculated phase overlap. In Fig. 7 the relative watts per inch for phase overlap is plotted along the billet length and is compared to that without phase overlap.

Appendix I. Calculation of Field Intensity Near End of Long Cylindrical Coil at Load Surface

Refer to Fig. 8 and the following assumptions:

1. Ratio of radial air gap to radius of load small enough such that curvature effects may be neglected.
2. Reference depth d small in relation to h and radiuses.
3. Constant coil ampere-turn per inch.

With the above assumptions the field intensity $H/(NI/L)$ is given by the following equation:

$$H/NI/L=1/2+1/\pi\tan^{-1}(x/h) \quad (3)$$

Equation 3 is plotted in Fig. 9 for values $+x/h$ and for values $(-x/h)$; see reference 1.

Appendix II. Effect of Total Field Intensity in Phase Overlapped Region

With reference to Fig. 10 where the vectors $\mathbf{H_1}$ and $\mathbf{H_2}$ represent the components of magnetic intensity at x produced by left and right hand coil segments and where the real axis is taken as the resultant of $\mathbf{H_1}+\mathbf{H_2}$ at $x=0$, and where H_1' and H_2' are the real components of $\mathbf{H_1}$ and $\mathbf{H_2}$ and H_1'' and H_2'' are imaginary components, the magnitude of the axial component of $\mathbf{H}$ at x is:

$$H=\sqrt{(H_1'+H_2')^2+(H_1''-H_2'')^2} \quad (4)$$

$$H=\sqrt{H_1^2+H_2^2+H_1H_2}$$

Substituting equation 3 into 4 for corresponding points and replacing

$1/2+1/\pi\tan^{-1}x/h$ by $f(x/h+\phi/2h)$

$1/2+1/\pi\tan^{-1}-x/h$ by $f(-x/h+\phi/2h)$

$$H^2/(NI/L)^2=f^2(x/h+\phi/2h)+f^2(-x/h+\phi/2h)+f(x/h+\phi/2h)f(-x/h+\phi/2h) \quad (5)$$

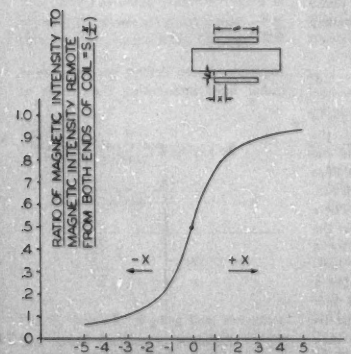
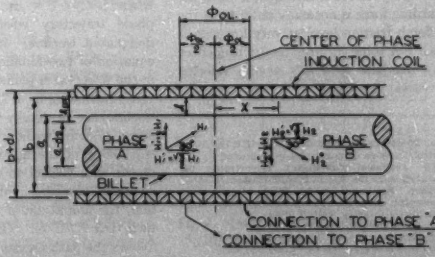


Fig. 9 (left). Ratio x/h of distance in coil to radial air gap

Fig. 10 (below). Cross section of coil and billet at an overlapped phase point



In equation 5 in order to have the same heating at $x=0$ as at $x=\infty\ \phi/2h$ must be chosen such that:

$$\int_0^{\infty} f^2(\phi/2h) = 1 \qquad (6)$$

Writing equation 6 in the form of equation 3 it follows that

$$(\phi/2h) = \sqrt{1/3} = 1/2 + 1/\pi \tan^{-1}(\phi/2h) \qquad (7)$$

Solving equation 7: $\phi/2h = 0.247$. Hence the phase overlap ϕ should be equal to 0.494 h or approximately half of the effective radial air gap. The effective radial air gap is defined by

$$h = (b+d_1/2) - (a-d_2/2)$$
$$h = 1/2[(b+d_1) - (a-d_2)]$$

Therefore, the phase overlap is given by the equation

$$\phi_{OS} = 1/4[(b+d_1) - (a-d_2)]$$

Reference

1. RADIO FREQUENCY HEATING, G. H. Brown, C. N. Hoyler, R. A. Bierwerth. D. Van Nostrand Company, Inc., Princeton, N. J., 1947, chap. 8, p. 90.

Stability Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction

M. P. PASTEL
NONMEMBER AIEE

G. J. THALER
MEMBER AIEE

IT IS WELL KNOWN that relative motion between rubbing surfaces is opposed by friction forces. In general, a force directly proportional to velocity is produced by a component known as "viscous friction." Normally, another component, the coulomb friction force C, produces a force that is essentially constant regardless of velocity. In addition, if the surfaces are initially stationary, the force required to start motion is greater than the coulomb friction force. Since the surfaces seem to adhere, the effect is called "stiction" and the magnitude of the required starting force is called the stiction force S. These effects are illustrated graphically in Fig. 1. Note that the ordinate may be force or torque, and the dimensions of C or S obviously must be adjusted as needed.

The effect known as stiction is frequently considered to be a static (zero velocity) phenomenon, which postulates that as soon as motion is initiated the stiction disappears and only the coulomb friction remains. This is a reasonably accurate assumption since the retarding force is normally reduced to the coulomb friction level at very low velocities. For some surfaces the stiction effect may decrease gradually as the velocity increases; see Fig. 1.

Paper 59-649, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959. Manuscript submitted February 10, 1959; made available for printing April 3, 1959.

M. P. PASTEL and G. J. THALER are with the U. S. Naval Postgraduate School, Monterey, Calif.

Phase Plane Analysis

Step Function Response of a Second-Order Servo

For a second-order proportional-error servomechanism with coulomb friction, the torque equilibrium equation is

$$J\ddot{\theta}_c + f\dot{\theta}_c + C \text{ sign } \dot{\theta}_c = KE \qquad (1)$$

where

J = polar moment of inertia
F = coefficient of viscous friction
C = coulomb friction torque
K = torque constant
θ_c = output position angle
E = error = $\theta_R - \theta_c$

For a step displacement input equation 1 becomes

$$J\ddot{E} + f\dot{E} + KE + C \text{ sign } \dot{E} = 0 \qquad (2)$$

To analyze the system on the phase plane the isocline equation is obtained from equation 2 and is

$$\dot{E} = -\frac{KE}{NJ+f} - \frac{C \text{ sign } \dot{E}}{NJ+f} \qquad (3)$$

where $0 < N < \mp\infty$ and N is the slope of phase trajectory where it crosses the designated isocline. Equation 3 is the equation of two families of straight lines in the $\dot{E}$ versus E plane, as shown in Fig. 2(A), emanating from $E = \mp C/K$. These isoclines are identically the isoclines of the linear system (without C) except for a translation of the upper and lower half-planes with respect to the origin. For a step displacement input to initiate motion, the amplitude of the step must be such that $E > C/K$. The trajectory then follows the path dictated by the isoclines until it terminates on the E-axis between the foci. If discontinuous stiction is present the step must be such that $E > S/K$ to initiate motion, the trajectory still follows the isoclines emanating from the foci at $\mp C/K$, but the trajectory terminates (system sticks) if it hits the E-axis between $\mp S/K$. This is illustrated in Fig. 2(A). A step displacement of magnitude such as to cause $E = A$ (in Fig. 2) would start a trajectory for coulomb friction without stiction, but it does not initiate motion when stiction is present. For a step large enough to set $E = A'$ motion is always initiated. The trajectory $A'BD$ is produced for C without S, but is terminated at B if stiction is present. Note that in either case $E_{ss} \neq 0$. Coulomb friction and stiction are always stabilizing when a step displacement input is used. This can be used to stabilize unstable systems partially, if some static error is acceptable.

Ramp Function Response of Second-Order System With Coulomb Friction Only

When the input is a suddenly applied ramp function and only coulomb friction is present, equation 1 still applies, and

$$J\ddot{E} + f\dot{E} + KE = f\omega_t + C \text{ sign } (\dot{\theta}_R - \dot{E}) \qquad (4)$$

where $\omega_t = \dot{\theta}_R =$ input velocity. The isocline equation becomes

$$\dot{E} = -\frac{KE}{NJ+f} + \frac{f\omega_t + C \text{ sign } (\dot{\theta}_R - \dot{E})}{NJ+f} \qquad (5)$$

Equation 5 also defines two families of radial lines intersecting the E-axis at

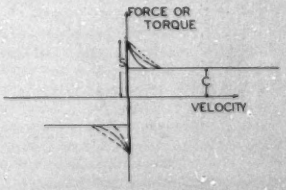


Fig. 1. Definition of coulomb friction and stiction

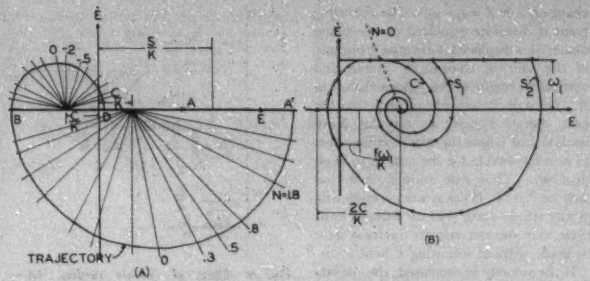


Fig. 2. A—Phase trajectory of second-order servo with coulomb friction and stiction. B—Phase trajectories as in A, plus ramp input

$E=(f\omega_t \mp C)/K$. The foci are again separated by 2 C/K, but the origin of co-ordinates is not midway between these foci; it is displaced from this mid-point by an amount $f\omega_t/K$. Also, the term $C(\text{sign } \dot{\theta}_R - \text{sign } \dot{E})$ designates the set of isoclines to be used, and thus defines a dividing line at $\dot{E}=\omega_t$. This is shown in Fig. 3 for an arbitrarily selected ω_t. The phase trajectory for a ramp input has initial conditions $E=0$, $\dot{E}=\omega_t$, and thus starts at point P. Motion is initiated when $KE=C$, corresponding to a horizontal displacement from point P of amount $E=C/K$. This is precisely the point at which the $N=0$ isocline crosses the dividing line. After one complete spiral about the focus, the trajectory must cross the $N=0$ isocline below the dividing line if the linear system has $\zeta>0$. The phase trajectory spirals into the focus and the steady-state condition provides a velocity lag error

$$E_{ss}=\frac{f\omega_t+C}{K} \qquad (6)$$

Note that the phase plane of Fig. 3 is adjusted to only one input velocity. For each new ω_t the dividing line and the origin must be shifted. The displacement of the origin from the point midway between the foci is $-(f\omega_t)/K$, and the vertical displacement of the dividing line is $+\omega_t$. Thus, as ω_t varies from zero to infinity, the intersection of the dividing line with the $\dot{E}$-axis moves along a line with slope $-K/f$. This is precisely the slope of the $N=0$ isocline, and the displacement between these lines is C/K. Thus for any ω_t the system is stable, if the linear system is stable.

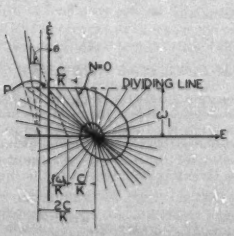


Fig. 3. Isoclines, dividing line, and phase trajectory of second-order servo with coulomb friction and ramp input: $\theta=\tan^{-1} K/f$

Ramp Function Response of a Second-Order System With Coulomb Friction and Discontinuous Stiction

When stiction is present but discontinuous (disappearing for $\dot{\theta}_R \neq 0$) the basic isocline configuration of Fig. 3 is unchanged, but motion is not initiated until the error is such that $KE=S$. This makes a limit cycle possible when a ramp function is suddenly applied, providing that ω_t is sufficiently low. Fig. 2(B) shows the phase plane for a ramp input ω_t, and 3-phase trajectories, for C only, for S_1 slightly greater than C, and for S_2 much greater than C. The first two are seen to be stable, but for S_2 the phase trajectory is a spiral of such dimensions that it intersects the dividing line, the output member sticks, and a limit cycle results.

Stability Criterion for a Second Order System With Discontinuous Stiction When a Ramp is Suddenly Applied

Knowing the system parameters, the values of C and of S, and the value of ω_t, makes it possible to predict whether a limit cycle will exist when the ramp is suddenly applied. Note that the phase trajectory is defined by the isoclines of the linear system, since the only effect of C is to translate the focus. Thus a linear equation can be derived for the phase portrait. Furthermore, the specific trajectory can be designated by inserting the co-ordinates of the starting point of the trajectory when the constant of integration is to be evaluated. The necessary condition for the existence of a limit cycle is apparent from the phase portrait: as the trajectory spirals around the focus the first time, it must pass through a maximum velocity; and if this value of $\dot{E}$ equals or exceeds ω_t, the system sticks and a limit cycle is started. To derive a stability criterion from these conditions, the equation for the phase trajectory is given by

$$\ddot{E}^2+2\zeta\omega_n E\dot{E}-4\zeta^2\omega_t\dot{E}+\omega_n^2E^2-4\zeta\omega_n\omega_t E+4\zeta^2\omega_t^2=$$

$$D_\theta \frac{2\zeta}{\sqrt{1-\zeta^2}} \tan^{-1}\frac{\omega_n E+\zeta\dot{E}-2\zeta\omega_t}{\dot{E}\sqrt{1-\zeta^2}} \qquad (7)$$

where $\zeta=f/2\sqrt{KJ}$; $\omega_n=\sqrt{K/J}$; $D=$ constant of integration. The specific phase trajectory may be defined by substituting the co-ordinates of the point at which the system pulls free of the stiction, evaluating $D=D_1$. Equation 7 is then completely defined, and it is possible to evaluate the maximum velocity reached on the first spiral. A simpler procedure is to note that the maximum velocity must be a point somewhere on the $N=0$ isocline. The critical phase trajectory is the one which passes through the point of intersection between the $N=0$ isocline and the dividing line. If the value of D associated with this critical trajectory is computed as D_2, then the system has a limit cycle if $D_1 \geqq D_2$. By proper manipulation the stability criteria is obtained as

$$[1 \geqq -b(b-2\zeta)]\, e^{\frac{a\tan^{-1}\frac{1-2\zeta^2}{-2\zeta a'}-\tan^{-1}\frac{1+\zeta(b-2\zeta)}{a'(b-2\zeta)}}} \qquad (8)$$

where a, a', and b are as defined in the Appendix. Fig. 4 presents the stability criterion as a curve.

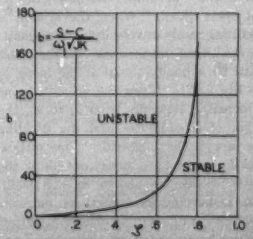


Fig. 4. Graph of stability criterion for second-order servos with stiction and ramp input

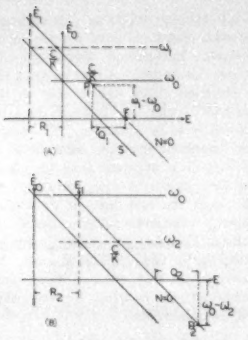


Fig. 5. Effect of changing ramp value

A—Velocity increase: $Q_1 = R_1 = 2\zeta(\omega - \omega_0)/\omega_n$. S = locus of points at which vertical axis intersects dividing line

B—Velocity decrease: $Q_2 = R_2 = 2\zeta(\omega_0 - \omega_2)/\omega_n$

Effect of a Change in Input Velocity: Stability Criteria

For any given system in which stiction is present a limit cycle must result if the system is started from rest with a sufficiently small ramp ω_t, and a limit cycle cannot result if ω_t is sufficiently large. If stiction produces limit cycles at useful speeds, the limit cycle can be avoided by starting with a large ω_t and reducing the input velocity slowly. If the reduction in speed is accomplished too rapidly the limit cycle may be initiated. Once the system has been adjusted to operate stably at a low speed, the speed may be increased as rapidly as desired without initiating a limit cycle, but speed decreases must be made slowly. This section presents an explanation of these phenomena, and develops certain stability criteria for the speed adjustments.

Fig. 5 shows the effect of changing ω_t on certain important characteristics of the phase plane. When the system is operating stably at velocity ω_0, the operating conditions are described on the phase plane by a point located at the focus F; see Fig. 5(A). When the velocity input is increased suddenly from ω_0 to ω_t, the origin of co-ordinates and the dividing line are both displaced. The critical sticking point C/K moves out along the $N=0$ isocline, as shown in Fig. 5(A). The instantaneous conditions may be described as follows: no change in output position or velocity; no change in input position; but a change $(\omega_t - \omega_0)$ in input velocity. The instantaneous error is unchanged, but $\dot{E} = \omega_t - \omega_0$. The operating point is therefore displaced on the phase plane; it is displaced horizontally because of the shift in co-ordinate origin, and vertically because of the sudden change in $\dot{E}$. These displacements are such that the new operating point is on the $N=0$ isocline and below the new dividing line, so it is impossible for the output to stick (if $\zeta > 0$). Thus, one stability criterion may be stated: if the system is operating at any steady-state velocity without limit cycle, any desired velocity increase may be made without initiating a limit cycle.

If the velocity is decreased, the operating point moves in a similar fashion but in the opposite direction to point P_2 in Fig. 5(B). Since P_2 and the critical point C/K are both on the $N=0$ isocline, and the limit of stability is reached if the trajectory from P_2 passes through C/K, then the system is stable if

$$\omega_2 > (\omega_0 - \omega_2)e^{-\pi\zeta} \tag{9}$$

from which

$$\omega_2 > \frac{\omega_0}{1 + e^{\pi\zeta}} \tag{10}$$

Note that equation 10 states the stability criterion for a sudden decrease in input velocity.

Effect of Variable Stiction

When the stiction-coulomb friction characteristic is variable rather than discontinuous, as shown by the broken lines in Fig. 1, the analysis of the preceding paragraphs is not complete. The stability criterion of Fig. 4 is only partially correct in the sense that if Fig. 4 denotes an unstable system it is surely unstable, but some systems which appear to be barely stable may actually be unstable if the stiction is variable rather than discontinuous. For variable stiction the location of the focal point on the E-axis is given by

$$E = \frac{2\zeta\omega_t}{\omega_n} + \frac{S_x}{K} \tag{11}$$

where S_x is the value of stiction specified by the instantaneous velocity. Thus the focal point moves along the E-axis as $\dot{E}$ changes. A step-by-step construction may be used to determine the phase trajectory. First, the stiction versus velocity characteristic must be known, and it is convenient to use a sliding overlay on a base drawing of the isoclines. As indicated in Fig. 6, the breakaway point is located at S_0/K, for which the focal point is at zero. An increment of the trajectory is constructed to point A. One then refers to the stiction versus velocity characteristic to evaluate S_x and equation 11 is used to locate a new focal point at a. The curve is then constructed from A to B, a new focal point located at b, the trajectory constructed from B to C, a new focal point located at C, etc. The trajectory for discontinuous stiction is also shown in Fig. 6. Since the focal point for this case is fixed at X, it is obvious that variable stiction has a greater destabilization effect than discontinuous stiction.

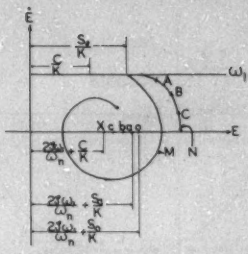


Fig. 6. Effect of variable stiction: M = trajectory for discontinuous stiction; N = trajectory for variable stiction

Conclusions

The effect of coulomb friction, discontinuous stiction, and variable stiction have been explained with regard to second-order servos. The conditions which lead to a limit cycle with ramp input have been determined and stability criteria have been developed for the permissible magnitude of a suddenly applied ramp and for the permissible variations of input velocity during operation.

Appendix. Derivation of Stability Criterion for Suddenly Applied Ramp Input

Definitions:

$$a = 2\zeta/\sqrt{1-\zeta^2},\ a' = \sqrt{1-\zeta^2},\ R = \omega_t/\omega_n$$

$$b = (S-C)/KR,\ E_r = E - (2\zeta R + C/K)$$

Equation 7 gives the phase trajectory for a linear system in response to a ramp input. If the origin of co-ordinates is shifted to the focal point so that the relative error quantity is designated E_r, then the equation of the phase trajectory with respect to this origin is

$$\dot{E}^2 + 2\zeta\omega_n\dot{E}E_r + \omega_n^2 E_r^2 = D_e^{\frac{a}{a'}\tan^{-1}\frac{\dot{E}+\zeta\omega_n E_r}{a'\omega_n E_r}}$$

To evaluate D_1 and D_2 for the phase trajectory from the break-free point and for the phase trajectory through the intersection

of the $N=0$ isocline with the dividing line, note that the co-ordinates of point 1 are $\dot{E}=\omega_t=R\omega_n$; $E_r=bR-2\zeta R$; and for point 2, $\dot{E}=R\omega_n$; $E_r=-2\zeta R$. Substituting these values yields

$$D_1=[R^2\omega_n^2+2\zeta\omega_n^2R^2(b-2\zeta)+\omega_n^2R^2(b-2\zeta)^2]e^{-a\tan^{-1}\frac{R\omega_n+\zeta^2\omega_nR(b-2\zeta)}{a'\omega_nR(b-2\zeta)}}$$

$$D_2=[R^2\omega_n^2-4\zeta^2\omega_n^2R^2+4\zeta^2\omega_n^2R^2]e^{-a\tan^{-1}\frac{R\omega_n-2\zeta^2\omega_nR}{-2a'\omega_nR}}$$

and

$$\frac{D_1}{D_2}=[1+b(b-2\zeta)]e^{a\left(\tan^{-1}\frac{1-2\zeta^2}{-2\zeta a'}-\tan^{-1}\frac{1+\zeta(b-2\zeta)}{a'(b-2\zeta)}\right)}$$

The system goes into a limit cycle if $D_1=D_2$; therefore the stability criterion is

$$1\leq[1+b(b-2\zeta)]e^{a\left(\tan^{-1}\frac{1-2\zeta^2}{-2\zeta a'}-\tan^{-1}\frac{1+\zeta(b-2\zeta)}{a'(b-2\zeta)}\right)}$$

Since the only variables are b and ζ, a curve may be obtained on the b versus ζ plane in such a way that all points on the curve satisfy the condition $D_1/D_2=1.0$, and this curve divides the plane into stable and unstable regions.

Reference

1. Operating Modes of a Servomechanism with Nonlinear Friction, H. Laner. *Journal*, Franklin Institute, Philadelphia, Pa., 1955.

Constant-Frequency Variable-Speed Frequency-Make-Up Generators

BERT V. HOARD
MEMBER AIEE

EXPERIENCE obtained on constant-frequency generation systems which are installed in jet aircraft has indicated areas where improvements are needed. Examples are:

1. Unpredictable failures, such as carbon brush dusting.
2. Relatively short life of parts subject to wear.
3. Poor quality of frequency supply.

Always present also are the needs for reductions in weight, volume, maintenance, and rebuilding costs; for improvement in efficiency; and for increased available time for operation.

Elimination of brushes on the generator and use of static controls, recently used on some jet aircraft, have approximately doubled the mean-time-between-failure of generator and drive unit. However, the drive has a number of wearing parts, and as the wear progresses a part may fail. Or due to increased hydraulic leakage the quality of frequency control deteriorates. Sufficient improvement in frequency control can eliminate special high-quality power supplies otherwise needed for sensitive loads. This also improves the performance of the electric system as a whole.

The Frequency-Make-Up Generator System

By replacement of the generator and mechanical constant-speed drive with a generator system designated hereafter as a frequency-make-up generator, it is estimated that the mean time between failure can be perhaps 4 times that of the orginal brush system or about twice that of the brushless system. The quality of frequency control will be improved and will not be subject to wear of parts. If in addition an improvement in bearing life and a decrease in unpredictable bearing failures can be attained, much better reliability will result.

It is predicted that weight, volume, efficiency, and maintenance for the new system will be either equal to or improved over the existing brushless generator with mechanical constant-speed drive.

Theory and Operation

The frequency-make-up generator has a standard 3-phase stator but has a wound-rotor type field. A 3-phase field will be used for descriptive purposes. The principle of operation used is to supply a correct value of a make-up or slip-frequency power to the polyphase wound field of the generator. By this means, the algebraic sum of; (a) the variable frequency provided from shaft rotational speed, and (b) the electrical make-up frequency introduced into the field windings; together, provide an air gap flux rotating with respect to the stator at constant speed. Under these conditions the stator output frequency is always constant.

A generator of this type may be designed to operate all below synchronous speed; all above synchronous speed; or below, through, and above synchronous speed.

Induction machine theory, for a generator of this type,[1] states in effect that if there is no shaft speed, the slip is 1.0 and all stator output power must be transferred from the rotor through the air gap by transformer action. At half synchronous speed the slip is 0.5, one half of the stator power is received by transformer action from the rotor and one half mechanically from shaft rotation. At synchronous speed the slip is zero, and all stator power is obtained from shaft rotation. Only d-c power is supplied to the rotor at synchronous speed for resistance losses, which of course are present at any speed.

At 1.5 times synchronous speed, the slip frequency in the rotor has reversed (-0.5 slip) the shaft is supplying 1.5 times stator output power, and the excess, 0.5 of the stator power, is flowing out of the rotor and must be recovered or the efficiency becomes very low. For all these cases, the stator frequency multiplied by the slip determines the required rotor frequency.

At a constant stator voltage, and a constant stator output power and power factor, but a varying shaft speed:

1. The shaft torque remains essentially constant, but shaft power input increases directly with shaft speed.
2. The rotor excitation current remains essentially constant.
3. Rotor input power increases directly with the slip below synchronous speed. I^2R rotor losses must be added.
4. Rotor output power increases directly with slip above synchronous speed. I^2R rotor losses reduce the rotor output.
5. Rotor voltage increases with slip either above or below synchronous speed to provide for increased reactance drop and increased power.
6. Friction, windage, and rotor iron losses modify these results slightly with change

Paper 59-773, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 16, 1959; made available for printing April 8, 1959.

Bert V. Hoard is with the Boeing Airplane Company, Seattle, Wash.

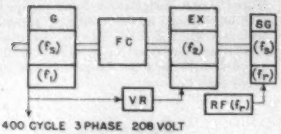


Fig. 1. Frequency-make-up generator with variable speed shaft

SG = signal generator
FC = frequency changer
RF = reference frequency
VR = voltage regulator
G = generator
Ex = exciter

in speed. Stator iron and copper losses are essentially constant.

For machines operating in parallel at any speed and constant total load:

1. An increase in rotor current magnitude will increase reactive power supplied to the system by a generator.

2. An advance in rotor current phase angle will increase real power supplied by a generator to the system. Below synchronous speed this requires more power input to the rotor but above synchronous speed this requires recovery of more power from the rotor.

Selection of a Generator System

The frequency-make-up system selected for modern aircraft should have no brushes and a minimum number of parts subject to wear. Weight and relative complexity become important comparison factors when considering different systems of this type.

The generator system shown in Fig. 1 will be described and discussed and then a weight comparison will be made. As shown, the generator is in its simplest form for operation all below synchronous speed. With added complication it can be made to operate all above synchronous speed or below through and above synchronous speed.

The system components shown in Fig. 1 consist of:

1. A generator having a 3-phase 400-cycle stator and a 3-phase wound rotating field. (Stator frequency $f_1 = 400$ cycles.)

2. An exciter with stationary d-c field and a 3-phase wound rotor. The rotor output supplies power at a variable frequency, f_2, equivalent to shaft speed.

3. A reference frequency source, f_r, 2-phase, for excitation supply to a signal generator. The main generator output frequency f_1 will equal f_r.

4. A signal generator with 2-phase stator and 3-phase rotor. The reference frequency f_r is supplied to the stator. The rotational speed of the shaft provides an equivalent f_2 frequency. The rotor output provides gating signals at slip frequency f_s.

5. A 3-phase frequency changer or rectifier-inverter made of static components (transistors or controlled rectifiers) but rotating to eliminate the need for brushes and commutator or slip rings. This frequency changer accepts variable frequency f_2 power from the exciter rotor and converts it into slip frequency f_s power for supply to the generator rotor. Slip frequency signals for frequency changer control are obtained from isolated windings on the signal generator rotor.

6. A voltage regulator which varies the d-c field excitation of the exciter, hence the magnitude of slip-frequency current in the generator field. This keeps the output voltage of the generator constant.

For parallel operation of generators, a phase shift in the reference frequency supply to the signal generator or in the signal generator stator windings will provide for proper sharing of the load between generators.

This generator, operating all below synchronous speed, will meet the following conditions:

1. Operates efficiently over an engine input shaft speed ratio of 0.5 minimum to 1.0 maximum. This speed range is compatible for most jet-type aircraft engines in use or planned.

2. In terms of synchronous speed the range will be 0.45 minimum and 0.9 maximum. For any speeds closer than 0.95 of synchronous speed the signal generator output signal becomes too low for reliable gating of the transistors or controlled diodes in the frequency changer.

3. Synchronous speed selected is 12,000 rpm, so maximum shaft speed at 0.9 synchronous speed will be 10,800 rpm.

4. It has minimum complexity for control of the frequency changer.

The generator operating through and above synchronous speed requires a basically different and more complicated method for obtaining gating signals. These can be obtained by adding and subtracting two equal magnitude voltages; see Appendix. One of these voltages varies at the reference or system frequency rate f_1, the other varying at the shaft frequency rate f_2. The resultant, when properly rectified and filtered, will provide slip-frequency signals for supplying exciter power modified to slip frequency to the generator rotor.

When operating above synchronous speed, the rotor output power must be efficiently recovered. This requires controlled inversion, in that the low slip frequency f_s wave must be chopped up to provide higher frequency f_2 for feedback into the exciter. When operating below synchronous speed, controlled rectification, a simpler process, is used.

Weight Comparison

In Table I there is shown, for a 60-kva generator, a weight comparison for an all-below system, *AB*, an all-above system, *AA*, and two below-through-above systems, *BTA*-8 for 8,000 rpm and *BTA*-6 for 6,000 rpm synchronous speeds respectively. In making these comparisons

Table I. Weight Comparison of 60-Kva Frequency-Make-Up Generators

	System Designation			
	AB	AA	BTA-8	BTA-6
	Generator Synchronous Speed, RPM			
	12,000	6,000	8,000	6,000
Per cent of minimum and maximum speed range	45 — 90	110 — 220	75 — 150	75 — 150
Speed range, minimum and maximum rpm	5,400 — 10,800	6,600 — 13,200	6,000 — 12,000	4,500 — 9,000
Generator rotor power (sK)	0.55K — 0.10K	−0.10K — −1.20K	0.25K — −0.5K	0.25K — −0.50K
Exciter Rating*				
Power out (1.08×1.06×sK)	0.63K		0.286K	0.286K
Power in (0.92×0.94×sK)		−1.04K	−0.433K	−0.433K
Relative weight: Generator	0.615	1.0	0.82 — 0.82	1.0 — 1.0
Relative weight: Exciter	0.79	0.592 — 0.42	0.342 — 0.51	0.42
Relative weight: Exciter and generator	1.395	1.592 — 1.24	1.162 — 1.51	1.42
Weight 60-kva exciter and generator†	126 lbs.	143 lbs. — 111.7 lbs.	136 lbs.	

* Includes 6% I^2R loss in generator rotor and 8% loss in frequency changer.

† Generator kva at synchronous rpm
Exciter kva at minimum or maximum rpm
Weight = 90[kva/60×6,000/rpm]$^{0.7}$ lbs
s = slip in per unit
K = generator rating kva

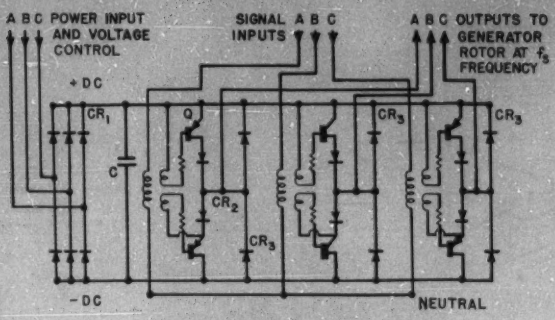


Fig. 2. Transistorized frequency changer

CR_1 = input rectifier
CR_2 = blocking rectifier
CR_3 = by-pass rectifier
Q = transistor
C = capacitor

only the two major weight components are used. These are the generator and the exciter. These weights change materially with assumed shaft speeds and generator ratings. Fixed weight components are not included as they are equal in all systems. Any weight comparison must be made on the basis of approximately equal maximum values of shaft speed and centrifugal force stresses in the generator or exciter rotor. Exciter weight is determined, except for the *AA* system, at minimum shaft speed when maximum output is required.

For 400-cycle generators the synchronous speeds available are limited to 6,000, 8,000, 12,000, or 24,000 rpm. Generators operating at speeds below 6,000 rpm will definitely be too heavy for a 60-kva size and a speed of 24,000 rpm probably is too high because of centrifugal stresses. With wound-rotor construction as compared to salient-pole construction, for a 60-kva generator, it is believed that lowest weight generators can be built with a maximum shaft speed in the neighborhood of 10,000 to 12,000 rpm.

In Table I the weight of the *AA* system is heavier and its maximum speed is higher than for any of the other systems. Also, for most aircraft flight requirements it will operate at high slip for most of the flight time, hence at relatively lower average efficiency. For these reasons subsequent comparisons will not include the *AA* system.

The weight of the *AB* system is lighter than the *BTA*-6 system and heavier than the *BTA*-8 system. This appears to depend mostly on maximum shaft operating speed plus the fact that only certain values for synchronous speed can be selected.

The power which the exciter and frequency changer must handle on the *BTA* systems is about 70% of that required in the *AB* system. The frequency changer will tend to be lighter because of less power handling requirements but may actually be heavier because of more complicated control and signal requirements, including the provision for absorbing power through the exciter when operating above synchronous speed. It is assumed here that the weights of frequency changer and controls will be essentially the same in both systems. However, there will be less complicated controls in the *AB* system. The average efficiency of the *AB* system will be somewhat better because more engine operating time will occur during flight when the slip is low and efficiency high.

Test Machine

A schematic of a 3-phase transistorized frequency changer, used in a small test generator of about 1 kva, is shown in Fig. 2. The generator was assembled from commercially available components and operated at 60 cycles output. The frequency changer did not rotate, which facilitated making test measurements. Limited tests were made to establish "proof of principle."

The frequency changer operates as follows. The 60-cycle power, equivalent to exciter power, is recitified by CR_1 input rectifiers and direct current and voltage are obtained. Rotor output from the signal generator supplies slip frequency signals f_s to the transistors Q through insulating signal transformers. In each phase the two transistors act like a double-throw switch which changes position each half cycle. The transistors operate in respective phases with a phase displacement of 120 degrees at the slip-frequency rate. By this means they supply pulsating direct current to the respective phases of the generator rotor at slip frequency. Rectifiers CR_2 are used to prevent possible failure of transistors due to excessive back voltage. Rectifiers CR_3 are used to make it easier for the transistors to turn off and allow the inverter to

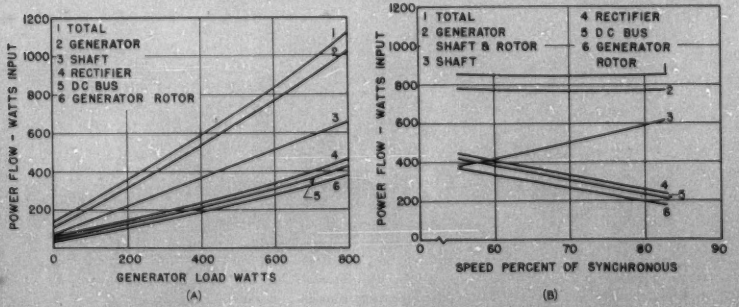


Fig. 3. Power inputs
A—For variable load at 70% synchronous speed
B—For variable speed at 600-watt load

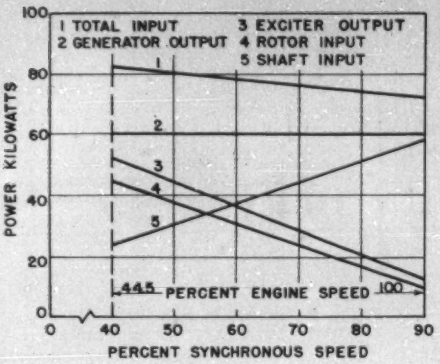


Fig. 4. Calculated power at 60-kw load speed range 40% to 90% of synchronous speed

carry reactive power in an efficient manner. For example, when transistor Q in rotor phase A turns off, the current in phase A rotor flows for a short additional period in the same direction, but through CR_3 rectifier of phase A. There is a slightly higher induced voltage, just enough to return the stored energy in the rotor phase back into the positive d-c through CR_3 rectifiers in the other phases.

Test Results

Power Relations

The curves in Fig. 3(A) show test generator power inputs for external loads from 0 to 800 watts, unity power factor, at 70% of synchronous speed. These curves show power to rectifiers, to the d-c bus, to the generator rotor, to the shaft, and total input over the load range. Shaft and rotor inputs practically straight lines over the range of generator loads shown. Higher test loads could not be carried because transistor temperature ratings were exceeded and a number failed.

The curves of Fig. 3(B) show test generator power inputs for an external load of 600 watts over a speed range from 55% to 85% of synchronous speed. Speeds closer to synchronous speed could not be obtained because of wrong taps on the signal transformers. Shaft speeds lower than those shown caused transistor temperatures to increase above rated values and again a number failed. In Fig. 3(B) it is of interest to note that, while total input (exciter losses excluded) remains nearly constant at any speed, the power supplied by the shaft decreases with speed and increased generator rotor power is required. If these curves are projected to 100% synchronous speed, only electric power input to supply d-c excitation losses would be required by the rotor; and shaft input would supply all other losses plus the load.

Harmonics and Unbalanced Loads

Harmonics in the output voltage are important but were not thoroughly investigated in the "proof of principle" tests. However, inspection of oscilloscope traces show generator stator voltages that appear close to sine waves.

It has been proposed that a floating synchronous condenser across the generator terminals will be needed to meet military specifications with respect to unbalanced voltages when unbalanced loads occur and to meet specification harmonic requirements. Unfortunately, in addition to the weight increase, two added bearings are introduced, that are subject to unpredictable failure, hence there will be reduced reliability and greater maintenance.

It is felt that all other possibilities should be investigated first, including a re-evaluation of military specification requirements, before a synchronous condenser having bearings is added to the system.

Semiconductor Failures and Ratings

The tests prove that if adequately rated semiconductors can be produced commercially, a frequency-make-up system with generator ratings up to 60 kva is practical. It is believed that this will occur in the near future. Germanium transistors were used in the tests and failures occurred, probably chiefly due to deterioration from junction temperatures higher than rated value but also from added effects of transient voltages and currents. Unlike the generator, transistors have little if any overload capability in current, voltage, or temperature, even for a few milliseconds above what is designated by one manufacturer as their absolute maximum ratings. Power-type controlled rectifiers were not available when the tests were started. Higher temperature silicon controlled rectifiers are now available in larger ratings than for transistors, and with some circuit modification they can be used. They have better overload characteristics than transistors. Before practical application can be made for generator service, the maximum voltage and current to which the semiconductors will be subjected for a given generator rating must first be determined. This will have to include generator overload, fault, overvoltage, and operational temperature conditions. In addition, for inverter or frequency-changer operation, repetitive peak voltages and currents which each semiconductor must handle are greater than rms rotor values indicate for any given loading condition on the generator. When these special conditions have all been satisfied to prevent failure during unusual operating conditions, then with normal loads on the semiconductors the percentage of their absolute maximum rating will be small, and very reliable operation should result.

Frequency Control

The frequency-make-up generator shows excellent frequency control when step loads are added or subtracted. When a step load is added, only a decrease in phase angle occurs because of the load addition. This phase angle decrease is recovered when the load is removed. These changes are much less than mechanical constant-speed drives can attain. If oven-controlled crystals are used as frequency reference, integrated error in frequency can be less than one cycle out of 10^5 cycles.

Calculated Performance, 60-Kva Generator

The curves of Fig. 4 show calculated input-output power flow curves for a 60-kva frequency-make-up generator operating at unity power factor. The curves are plotted with per-cent synchronous speed as abscissa, and for operation at full output from 40% to 90% of synchronous speed. The efficiency of this generator is several points better than can be achieved by a constant-speed drive and generator system in the important speed range from 60% to 90% of synchronous speed, where most in-flight operation usually occurs.

Conclusions

The following conclusions are obtained from the evaluation of frequency make-up systems studied, and their comparison with mechanical constant-speed drive and generator systems.

1. The *AB* synchronous speed system is preferred over a *BTA* system for a speed range of two to one or less because of less complicated signals, and controls. The weights of the two systems are estimated to be essentially the same, provided that equal maximum rotor speeds are used. The average operational efficiency of the *AB* system is slightly better than for the *BTA* system. Due to less complication, maintenance will be lower in the *AB* system. The rating of the frequency changer on the *BTA* system is about 70% of the rating of the frequency changer in the *AB* system.

2. If greater speed ranges than two to one are required, the weight of the *AB* system begins to exceed that of the *BTA* system and the *BTA* system might therefore be selected, although reliability would still be a major consideration.

3. Higher temperature and higher voltage and current ratings in controlled diodes and/or transistors are needed to build large class *C*-type generators at this time. Development of these semiconductors is progressing rapidly. It is probable, if development of such a generator is started at this time, that adequate semiconductors will become available by the time other generator development problems have been solved.

4. When comparing a 60-kva *AB* frequency-make-up system with a mechanical constant-speed drive and generator system, each having two-to-one speed ratio:

(a) The *AB* system weight is estimated to be 15 to 20% less.

(b) The volume and total length will be less.

(c) The efficiency will be about 3 to 5 points higher at cruise and maximum engine speeds.

(d) Maintenance and rebuilding costs will be materially less.

(e) Frequency deviation for step load changes will be reduced essentially to zero.

(f) In-flight failures will be reduced and reliability improved.

Appendix

To obtain a slip frequency signal through synchronous speed:

Using Sum and Difference Relationships[3] for two voltage waves e_1 and e_2, of equal peak magnitude E_0, but of different frequencies f_1 and f_2,

$$e_1 = E_m \sin 2\pi f_1 t$$

$$e_2 = E_m \sin 2\pi f_2 t$$

$$e_1 - e_2 = 2E_0 \sin \pi t(f_1 - f_2) \cos \pi t(f_1 + f_2)$$

$$e_1 + e_2 = 2E_0 \cos \pi t(f_1 - f_2) \sin \pi t(f_1 + f_2)$$

If $e_1 - e_2$ is rectified full wave and $e_1 + e_2$ is similarly rectified; and the results are subtracted and filtered; a resultant fundamental frequency $f_1 - f_2 = f_s$ is obtained where f_s equals slip frequency. Several high frequencies are also present before they are filtered out. This process is not efficient since a considerable portion of the power is lost in the transformation and filtering. If half wave rectification of the sum and difference equations is used, the proportion of signal output over input power is less than for full wave rectification. It will be observed that the magnitude of either the sum or the difference equation is identical, only the phase position is different.

Another area of investigation for obtaining a slip frequency signal through synchronous speed is suggested in the following manner:

If two voltages e_{a1} and e_{b1} can be made available by electronic means; for example, by modulation, variable impedance, etc. They are expressed as

$$e_{a1} = kf_1 I_0 \sin 2\pi f_1 t \sin 2\pi f_s t$$

$$e_{b1} = kf_1 I_0 \cos 2\pi f_1 t \cos 2\pi f_s t$$

$$e_{a1} = \frac{kf_1 I_0}{2} \cos 2\pi t(f_1 - f_2) - \cos 2\pi t(f_1 + f_2)$$

$$e_{b1} = \frac{kf_1 I_0}{2} \cos 2\pi t(f_1 - f_2) + \cos 2\pi t(f_1 + f_2)$$

$$e_{a1} + e_{b1} = kf_1 I_0 \cos 2\pi f_s t$$

Since

$f_s = f_1 - f_2$ is the needed slip frequency.

References

1. Alternating Current Machines (book), T. C. McFarland, D. Van Nostrand Company, Inc., Princeton, N. J., "Polyphase Induction Machines," pp. 143–216; "Impulse Method of Control of Grids," p. 71, p. 432; "Inverters," pp. 457–61, 1948.

2. Fundamentals of Vacuum Tubes (book), A. V. Eastman, McGraw-Hill Book Company, Inc., New York, N. Y., "Modulators and Demodulators," chaps. 13 and 14, 1949.

3. Mercury-Arc Power Converters in North America, AIEE Committee Report, *AIEE Transactions*, vol. 67, pt. II, 1948, pp. 1031–59.

4. Ignitor Excitation Circuits and Misfire Indication Circuits, A. H. Mittag, A. Schmidt, Jr. *Ibid.*, vol. 61, Aug. 1942, pp. 574–77.

5. Grid Controlled Rectifiers and Inverters, C. C. Hershind, *Electrical Engineering*, June 1934, pp. 926–35.

6. Design of an Electronic Frequency Changer, C. H. Willis, R. W. Kuenning, E. F. Christensen, B. D. Bedford, *AIEE Transactions*, vol. 63, 1944, pp. 1070–78.

7. The Electronic Converter for Exchange of Power, F. W. Cramer, L. W. Mortin, A. G. Darling, *Ibid.*, pp. 1059–69.

8. Analysis of Rectifier Circuits, E. F. Christensen, C. H. Willis, C. C. Hershind, *Ibid.*, pp. 1048–58.

9. Harmonics in the A-C Circuits of Grid-Controlled Rectifiers and Inverters, R. D. Evans, H. N. Muller, Jr. *Ibid.*, vol. 58, 1939, pp. 881–70.

10. History and Development of the Electronic Power Converter, E. F. W. Alexanderson, E. L. Phillipi, *Ibid.*, vol. 63, Sept. 1944, pp. 654–57.

11. Static Converter with Current and Voltage Smoothers for Flexibly Coupling a Three-Phase 50 Cycle Network with a Single Phase 16⅔ Cycle Network, C. Ehrensperger, *Brown Boveri Review*, Baden, Switzerland, June 1934, pp. 96–112.

12. The Flexible Coupling of Independent Three-Phase Networks by Means of a Single Mutator, E. Kern, *Ibid.*, Dec. 1934, pp. 214–18.

13. Synchronizing of an Induction Motor, E. McP Leyton, *U. S. Patent no. 2,650,335*, Aug. 25, 1953.

14. Electrical Apparatus, C. I. MacNeil, *U. S. Patent no. 2,959,044*, Nov. 10, 1953.

15. Constant Frequency Power Supply, W. E. Sargeant, *U. S. Patent no. 2,776,379*, Jan. 1, 1957.

16. Frequency Control Apparatus for Alternators, L. J. Johnsen, *U. S. Patent no. 2,854,617*, Sept. 30, 1958.

Discussion

K. M. Chirgwin and **L. J. Stratton** (Jack & Heintz, Inc., Cleveland, Ohio): It seems to us that it is the development of semiconductor frequency changers that has made vscf (variable speed constant frequency) generating systems for aircraft feasible at this time. It is appropriate, therefore, that Mr. Hoard's paper should have been the first to be read at the first Air Transportation Session on vscf generators, because Mr. Hoard was one of the pioneers of vscf generators which use semiconductor frequency changers.

There are a number of points in the paper which we would like to discuss. The first concerns Table I wherein the weights of the main rotating machines of four different arrangements are compared. In Table I the speed range is 2:1 for all of the four arrangements but the actual minimum and maximum speeds differ and, as Mr. Hoard points out, at least part of the weight differences shown in the table can be ascribed to these speed differences.

In Fig. 5 we have attempted to present the variables in such a way that these speed differences can be taken into account in making weight comparisons.

In this figure the weight of the generator and the weight of generator and exciter are both plotted as continuous functions of the speed which is selected as the synchronous speed of the generator. Some of the discrete synchronous speeds which can be selected to give 400-cps output frequency from the main generator are shown on the horizontal axis. Curves are shown for the same minimum speeds and for the same 2:1 speed range used by Mr. Hoard in his Table I. The curves are based on the assumptions that the weight of a rotating machine is proportioned to its kva divided by its speed of rotation, and that the exciter rating is the system kva times the maximum per-unit slip, i.e., the losses and magnetizing current of the system are neglected.

With these simplifying assumptions, the weight of the generator is inversely proportional to the synchronous speed selected and is independent of minimum speed and of speed range, as shown by the lower curve

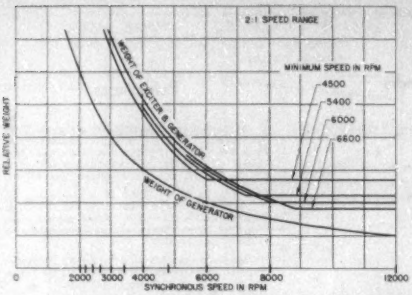


Fig. 5. Rotating machine weights versus the selected synchronous speed of the main generator

of Fig. 5. When the synchronous speed of the generator is selected below the minimum speed or in the lower part of the speed range, then the weight of the exciter is determined by the maximum speed and the slip at maximum speed. This condition gives the curved lines for the combined weight of exciter and generator shown in the left part of Fig. 5. When the synchronous speed of the generator is selected above a certain speed, "corner speed," in the lower part of the speed range, then the exciter weight is determined by the minimum speed and the slip at minimum speed. This condition results in the horizontal straight lines of the combined weight of exciter and generator shown in the right part of Fig. 5. The derivation of these curves is given in more detail in the Appendix.

The significance of these horizontal straight lines is that in this region the combined weight of the two rotating machines is independent of the synchronous speed selected and is inversely proportional to the minimum speed. It is the division of weight between the generator and the exciter which is determined by the synchronous speed selected. It should also be noted that the horizontal line is the lowest combined weight for a given minimum speed.

For an *AA* system the synchronous speed always falls in the curved combined weight region and is therefore always heavier than either a *BTA* or an *AB* system. This confirms Mr. Hoard's findings.

Provided the synchronous speed of the *BTA* system is selected above the "corner speed" then, with the simplifying assumptions, the combined weight is exactly the same as that of the *AB* system no matter what synchronous speed is selected.

Provided the synchronous speed of the *BTA* system can be selected near the middle of the speed range, then the power rating, and the weight and volume of the power handling components of the frequency changer are all about 50% of those of an *AB* system having the same minimum speed. Because of this and because the machine weights are equal, the *BTA* system is the lighter.

Mr. Hoard expects that there will be difficulties in operating the frequency changer with reversed power flow, and in maintaining control while passing through synchronous speed with a *BTA* system. We have found that these difficulties can be overcome with very little addition to the weight and volume of the frequency changer.

When we come to actual systems, generalizations are difficult to make because many other factors such as losses, magnetizing current, power factor, overloads, and fault capability must be included when making weight comparisons. We have found from detailed comparisons of systems for a number of different specifications that with the *BTA* system the volume of the frequency changer permits it to be located inside the shaft and that the *BTA* system has the lowest weight.

Since it is our company which has advocated the use of a floating synchronous machine which we term a Voltage Balancer, we feel it incumbent upon us to say a little about the need for such a machine. If the Voltage Balancer were omitted, then the requirements of the system in respect to reactive kva, waveform, and fault capability could all be met if suitable changes were made in other components of the system. But we believe either this machine or an equal induction machine is presently needed if the usual military specifications on voltage unbalance are to be met. Furthermore, we feel that considerable development must occur before this machine can be replaced by static components or eliminated altogether.

As will appear in the reply to a discussion of our paper[1] by J. T. Duane, the presence of the voltage balancer allows a reduction in the weight of the main generator and exciter, since it supplies reactive kva to the system. We believe Mr. Hoard takes an unduly pessimistic view of the bearing problems of this machine, which will likely be located in a more favorable environment than the main machine.

Appendix. Derivation of the Weight Comparison Curves

This derivation is based on two simplifying assumptions: 1. The weight of a rotating machine is proportional to its kva output divided by its speed of rotation; 2. The rating of the exciter is the system kva multiplied by the per-unit slip.

The main generator is a synchronous flux machine so its weight is proportional to the system kva K divided by the synchronous speed, N_s, selected for this machine.

$$\text{Generator weight} = C \cdot \frac{K}{N_s} \tag{1}$$

where C is the constant of proportionality. Plotting this equation on Fig. 5 gives a rectangular hyperbola, which is labelled Generator Weight.

If the variable speed shaft which drives the rotating machines is rotating at a speed N then the slip at which the synchronous flux generator is operating is given by

$$\text{Per-unit slip} = \frac{N_s - N}{N_s} \tag{2}$$

This equation results in negative values of slip when the shaft speed is higher than the synchronous speed of the generator.

The kva required of the exciter at this shaft speed N is

$$\text{Exciter kva} = K \cdot \frac{N_s - N}{N_s} \tag{3}$$

This equation gives negative values for the kva when N is above synchronous speed; the interpretation of this negative sign is that at these speeds the exciter is working as a synchronous motor and returning power to the shaft.

At this same shaft speed N the weight of the exciter required is given by

$$\text{Weight of exciter} = C \cdot \frac{K}{N} \cdot \frac{N_s - N}{N_s} \tag{4}$$

In this equation the negative sign that appears when the speed N is greater than synchronous speed should be ignored, since obviously the weight of the exciter when operating as a synchronous motor is not negative.

Adding equations 1 and 4 gives the combined weight of exciter and generator needed for the particular shaft speed N:

$$\text{Combined weight} = C \cdot \frac{K}{N_s} + C \cdot \frac{K}{N} \cdot \frac{N_s - N}{N_s} \tag{5A}$$

for speed N when $N < N_s$

$$= C \cdot \frac{K}{N_s} + C \cdot \frac{K}{N} \cdot \frac{N - N_s}{N_s} \tag{5B}$$

when $N > N_s$

Collecting like terms in equations 5 gives

$$\text{Combined weight} = C \cdot \frac{K}{N} \quad \text{when } N < N_s \tag{6A}$$

$$= 2C \cdot \frac{K}{N_s} - C \cdot \frac{K}{N} \quad \text{when } N > N_s \tag{6B}$$

In the vscf system the shaft speed N will vary over a speed range from $N_{min.}$ to $N_{max.}$ and machines of sufficient rating and weight must be provided to enable the system to put out the rated kva, K, at all speeds within this range. If N_s is selected so that $N_{max.} < N_s$, AB system, then equation 6(A) applies and the combined weight of the machines required for the speed range is

$$\text{Combined weight for speed range} = C \cdot \frac{K}{N_{min.}} \tag{7}$$

If N_s is selected so that $N_{min.} > N_s$, AA system, then equation 6(B) applies and the

combined weight of the machines required for the speed range is

Combined weight for speed range

$$=2C\cdot\frac{K}{N_s}-C\,\frac{K}{N_{max.}} \quad (8)$$

For the *BTA* system where $N_{min.}<N_s<N_{max.}$ it is not immediately apparent whether equation 6(A) or 6(B) applies. So both should be tried and the one that leads to the greater weight should be used. Doing this leads to the conclusion that equation 6(A) governs when N_s is in the upper part of the speed range and equation 7 gives the correct combined weight, that equation 6(B) governs when N_s is in the lower part of the speed range and equation 8 gives the correct combined weight, and that the change-over point occurs when the synchronous speed is so selected that the weight given by the two equations is equal; that is, when

$$C\cdot\frac{K}{N_{min.}}=2C\cdot\frac{K}{N_s}-C\cdot\frac{K}{N_{max.}}$$

or when

$$N_s=1/2\cdot\frac{N_{max.}\cdot N_{min.}}{N_{max.}+N_{min.}} \quad (9)$$

This value of the selected synchronous speed given by equation 9 is called the corner speed. Equation 7 therefore applies at all speeds above the corner speed and equation 8 at all speeds below the corner speed.

In Fig. 5 the weights given by equations 7 and 8 are plotted for the four values of $N_{min.}$ that Mr. Hoard uses in his Table I.

Reference

1. Variable-Speed Constant-Frequency Generator System for Aircraft, **K. M. Chirgwin, L. J. Stratton.** *AIEE Transactions*, pp. 304–10 of this issue.

D. S. Toffolo (U. S. Naval Research Laboratory, Washington, D. C.): The author speaking from experience, cites examples of areas where improvement of present day constant frequency systems is desired. The third example is a qualitative statement for, until the author defines poor, fair, good, and excellent quality of frequency supply in quantitative terms, his assessment of the situation remains subjective.

His statement concerning the elimination of special high quality power supplies needed for sensitive loads is somewhat gratuitous. For such sensitive loads are usually equally harmonic conscious if not more so. Yet the author recommends, if all else fails, a re-evaluation of military specification requirements.

To the remarks concerning parallel operation, the writer notes that this is what present systems do and in addition to which present parallel systems have synchronizing torque going for them. The author should realize that his and all other constant velocity magnetic field devices where the magnetic field has no fixed space relation to any set of windings in the generator, have no synchronizing torque between them. As a consequence, there will be serious oscillatory load division problems which the author will dearly wish he had never encountered. People who wander with frequency-make-up generators into the parallel system area where numerous are the pitfalls and deep the holes would be exceedingly wise to emulate the angels!

After reading the author's description of the system components and looking at the block diagram shown in Fig. 1, the writer is not convinced that proper real load division will occur simply because the author states that a phase shift in the reference frequency supply to the signal generator or in the signal generator stator windings will do the job. Has the author calculated the magnitude of the phase shifts required and the response times involved and how close to parity of load sharing such a scheme will achieve?

Further, as the writer has pointed out in his comments to another frequency-make-up generator proposal,[1] has the author considered what happens to his parallel system, in fact, to his entire generating capacity if the reference frequency is lost? Does the author not consider that such a failure is of the nature of an unpredictable failure and for him it is not only that but catastrophic as well?

In view of the remarks made by the author concerning re-evaluation of military specifications, the writer regards the weight comparisons developed by the author as irrelevant.

In the paper, the author states under "Test Machine" that limited tests were made to establish "proof of principle." Proof of what principle? Proof that the frequency-make-up principle is a unity power factor principle? Even that he has not proved. For all unity power factor loads are not linear. In fact, for some electronic loads, the flow of armature current is of the fractional cycle type, continuously, and such loads, completely 3-phase balanced, will play havoc with his generator. One could, of course, re-evaluate the military specifications again.

Under "Frequency Control" the author states that when a step load is added, only a decrease in phase angle occurs due to the load addition. The author does not define what this phase angle is. If he means by this the angle between the phasor representing the excitation voltage and the phasor representing the terminal voltage, his statement is incorrect. For a frequency-make-up generator, the power angle (the difference between the two phasors above) increases with load as in conventional generators.

In this writer's opinion, the remarks under "Conclusions" are not conclusions, but speculations. For instance, the circuit shown in Fig. 2 cannot be used for controlled rectifiers if one merely replaces the transistors Q by controlled rectifiers. Since the author does not include any other schematic diagrams, the writer assumes from the author's statements about simplicity et al, that it is his intention to do so. The author should investigate the problem of turning off the controlled rectifier once it has been turned on and tied to a positive d-c bus.

The writer has no quarrel with those who strive to develop constant frequency generators for flight vehicle systems. In fact, he is one of them. He does have a quarrel with those who presume to develop such a machine for flight vehicle conditions and then propose to change the system conditions and/or present their case, based on infinitesimal amounts of experimental data, in a rather speculative manner.

Reference

1. Precision Power Frequency with Variable Speed Generators, **S. E. Rauch, L. J. Johnson.** *AIEE CP 59-772* (available on request).

Bert V. Hoard: I appreciate the discussion of my paper by Mr. Chirgwin and Mr. Stratton very much. It is stimulating to have honest differences of opinion as to the best way to develop new types of apparatus because it usually results in better equipement. Actually, in this case our opinions are not far apart.

I believe that evaluation of weight and performance of the final hardware and presently unknown developments in the semiconductor field may well be the only means of determining which system will be better or preferable.

I think we can agree that if a rotating machine providing balancing action is necessary, and it could well be under present military specification requirements, that it is preferable to have a synchronous type balancer rather than an induction machine; that considerable reactive kva for the system can be advantageously supplied by this machine using d-c excitation rather than supplying it via the main exciter, the frequency changer, and the main generator; that weight and total generation losses could be saved.

Mr. Toffolo has placed his own interpretations on a number of the statements made in my paper. With most of these interpretations and conclusions I do not agree.

With present efforts to obtain nearly balanced loads on a-c aircraft systems, I believe that a re-evaluation of the minimum unbalanced voltage requirements for unbalanced loads called for in military specifications should be made, and if these requirements can be relaxed, some weight will be saved in any a-c generator with negligible decrease in performance as long as the load is balanced.

As to parallel operation of frequency-make-up generators, I foresee no paralleling problems anywhere near as difficult as those originally overcome before the 400-cycle a-c electrical system was finally successfully paralleled. An advance in the phase angle of the signal generator will gate excitation currents sooner in each cycle into the specific generator rotor and advances its phase angle electrically compared to another paralleled machine, and the machine will pick up a greater percentage of the total load. The machines have synchronizing torque as long as they have correct excitation. If a differential real load loop is used for controlling the relative phase angle of gating excitation currents into the various generator rotors, I am positive we can overcome Mr. Toffolo's suggested pitfalls, holes, and even the oscillations without special help from the angels. The response times for load equalization will be much faster than presently attainable by having to accelerate mechanical inertias to obtain phase angle shifts and provide for frequency recovery.

I do agree with Mr. Toffolo that this system, if it has a single frequency reference, is vulnerable to failure similar to the one he referenced. This does not mean that modifications to a single reference system cannot be made so that its failure will not be catastrophic to the system as a whole.

Regarding my statement "only a decrease in phase angle occurs due to addition of a step load," since the intended conclusions were not obtained by the reader, I should have amplified the statement. When a conventional generator which is driven by a constant speed drive has a large step load added to it, the governor and other controls act relatively slowly, and it is possible to lose more than a complete cycle before the frequency is restored to normal value. If the constant speed drive is badly worn and the load increment is large, more than one cycle may be lost. On the generator with the step load addition, there also occurs a phase angle lag in terminal voltage behind its no load phase angle with respect to air gap flux as a reference, since it is the cutting of air gap flux that produces machines internal voltage.

On the frequency-make-up generator with addition of an equivalent step load a phase angle shift only occurs. This phase angle shift might be assumed to be greater than for the conventional generator especially at low shaft speeds because of the increased lag between the applied excitation voltage and the resultant excitation current. Note that it is excitation current that produces air gap flux in this generator system. Actually with the effect of the relatively slow mechanical control action included, the total phase angle lag will be much greater in the conventional system. With a good reference system the frequency-make-up system may give integrated frequency accuracy of one part in 10^8 cycles and instantaneous frequency deviations considerably less than one cycle. Mechanical, frequency controlled systems do not approach these values.

I feel the weight comparisons for exciter and generator, the two heaviest pieces, are decidedly relevant for the comparison purposes used in the paper. If requirements are placed on one generator which increase its weight then a proportional weight increase will occur in the other generator system.

Variable-Speed Constant-Frequency Generator System for Aircraft

K. M. CHIRGWIN
ASSOCIATE MEMBER AIEE

L. J. STRATTON
ASSOCIATE MEMBER AIEE

Synopsis: Variable-speed constant-frequency (vscf) generator systems are studied as an extension of well-known vscf motor technology. The vscf generator systems become practical for aircraft main generating systems when the commutator function is performed by semiconductor devices. It is concluded that the best scheme for this application is one using as the main generator a wound rotor induction machine operated above and below synchronous speed, and a slip channel comprised of a semiconductor frequency changer and a synchronous machine driven by the variable-speed shaft.

This system was built and tested in the laboratory using commercial-type machines. Test results and oscillograms are presented.

EVER SINCE 400-cycle electric systems have been used on aircraft, it has been a goal of the aircraft accessory industry to generate this power from the variable-speed shaft of the prime mover without the need for a constant-speed drive. This is now possible using a vscf generator system.

A vscf generator system is here defined as an assemblage of rotating electric machines and other electric equipment that is capable of taking mechanical energy directly from a shaft which rotates at a variable speed and converting this energy into a-c electric energy at a constant frequency.

The vscf systems will likely have many applications. This paper is confined to the use of such a system as the main electric generating system of an aircraft. Application in this manner sets definite bounds to the system. For example, ratings are probably from 20 to 120 kva, the output is 120/208 volts, 3-phase, 400 cps (cycles per second), to meet specifications indicated in reference 1, or some similar specification. The input shaft rotates at conventional drive-pad speeds; that is, a speed range of between 2:1 and 3:1 with a mid-range speed of approximately 6,000 or 8,000 rpm.

VSCF System Classification

The vscf systems operated as motors[2] have been known for a long while, and a great deal of knowledge has accumulated over the years. In a vscf motor system the frequency is determined by the frequency of the synchronous supply system from which the motor derives its power. In a vscf generator system, however, a frequency reference signal must be added to the system in such a way that it determines the frequency at which the power is generated. Ways in which the frequency-reference signal may be integrated with the rest of the generating system will not be discussed, since they are outside the scope of this paper.

When the power-handling parts of the system are considered, the vscf motor and the vscf generator systems are essentially the same and can be conveniently classified in three groups as shown in Fig. 1 and Table 1. Fig. 1(A) shows a wound-rotor induction machine with a variable resistance connected across the slip rings. When operated as a motor on a constant frequency supply, the motor speed can be reduced below synchronous speed by an amount which depends upon the load on the shaft and the amount of resistance in the rotor circuit.[3] When operated as a vscf generator, this scheme allows the speed of the shaft to be raised above synchronous speed by an amount which is determined by the power output of the generator and the amount of resistance in the rotor circuit. For either the motor or the generator, power equal to the slip times the power output is dissipated in the resistances. Slip is defined by equation 1:

$$\text{Slip} = \frac{\text{Synchronous Speed} - \text{Actual Speed}}{\text{Synchronous Speed}} \qquad (1)$$

Table I. VSCF Power Circuits

Classification	Power Relations	Commutator Machines
Synchronous flux, wasteful, Fig. 1(A)*	slip power wasted in resistances	none
Synchronous flux, differential, Fig. 1(B)†	slip power returned to or supplied from shaft	handles slip power at slip frequency
Full rated, Fig. 1(C)	full power	handles full output power at 400-cps frequency

* Eliminated because losses are large in wide speed range.
† Single or double range. Double range halves slip, slip frequency, and slip power, but passing through synchronous speed is difficult.

Paper 59-872, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing May 4, 1959.

K. M. CHIRGWIN and L. J. STRATTON are with Jack and Heintz, Inc. Cleveland, Ohio.

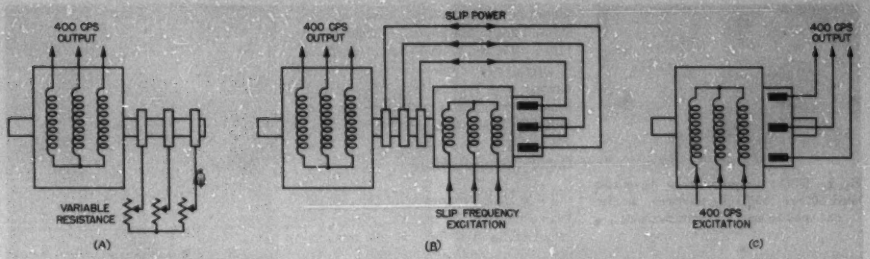


Fig. 1. Some vscf power circuits. A—Wound-rotor induction machine with resistance in the rotor circuit. B—Wound-rotor induction machine with Scherbius machine handling the slip power. C—Full rated scheme, shunt commutator machine

When the speed range is as wide as 2:1 or 3:1, the power lost in the resistances is large, being at least equal to the power output for a 2:1 speed range and at least twice the power output for a 3:1 speed range. This large amount of power has to be dissipated and, in addition, the airplane has to carry extra fuel in order to generate this additional power. Because of the penalties imposed by these two factors, wasteful schemes of this nature are not considered for aircraft application. In Table I this scheme is classified as a synchronous-flux system,[4] since regardless of the speed of rotation of the shaft, the air-gap flux of the induction machine rotates at synchronous speed.

Fig. 1(B), which is also classified as a synchronous-flux system, uses the same wound-rotor induction machine, but here, instead of slip power being wasted in resistances, it is returned to the shaft by means of an auxiliary commutator machine.

In the third classification, Fig. 1(C), the auxiliary commutator machine is increased in size to handle the full output power of the system and the induction machine is eliminated. This scheme is also a synchronous-flux scheme; however, schemes derived from this but which use other means of commutation do not have a synchronous flux. This scheme, therefore, is classified as a full-rated scheme because the commutator has to handle full output power of the system.

The arrangements shown in Figs. 1(B) and 1(C) are representative only. The vscf motor technology includes so many different arrangements that it is not possible to discuss them in detail. It is, however, possible to draw a number of general conclusions.

Included in the differential synchronous-flux category shown in Fig. 1(B) are slip power handling schemes which allow the system to operate on one side of synchronous speed only, because real power can flow in only one direction through one or more of the slip channel components. Such a slip channel cannot transmit reactive power because reactive power requires reversal of power flow once each cycle, so the system must be magnetized from the stator (high frequency) side of the induction machine.

There are other schemes in this category in which the slip channel can handle power in either direction, so the system can be magnetized from the rotor (low frequency) side of the induction machine, but the system is still confined to operation on one side of synchronous speed because, for one of a number of reasons, the system cannot pass through synchronous speed.

Table II. Commutating Components

	Two-Way Conduction Controllable	Uncontrolled	Rectifier Limited Control	Controllable
Components	rotating machine with commutator; mechanical rectifier; symmetrical transistor	uncontrolled rectifier (diode)	diode with modulator	controlled rectifier
Characteristics	any delay controlled rectification and inversion	no delay natural commutation, d-c only	limited delay controlled rectification	any delay controlled rectification and inversion
Application to vscf systems	eliminated	not possible	single range full rated; no lagging power-factor load	single range double range full rated; any power-factor load

Table III. Comparison of VSCF Systems

Items	Differential Synchronous Flux: Double Range	Differential Synchronous Flux: Single Range	Full Rated
Power in commutating components 2:1 speed range	1/3 of output kva	1/2 of output kva	full output kva
Pad speeds	compatible with present pad speeds		smaller ratings could use higher speeds, but gears needed with present engines
Total system weight	lightest	intermediate	heaviest
Efficiency	highest	lower	lower
Voltage balance	rotating voltage balancer		could use individual phase regulation
Parallel operation	no problem		individual phase regulation complicates problem
Wave form	combination of induction generator and rotating voltage balancer give good wave form		wave shaping necessary
Fault capability	inherent overload capabilities of rectifiers and machines are sufficient for usual faults		rectifiers must be increased in size for 5-per-unit faults
Complexity	similar to rotating rectifier machine with control of rectifiers (partly rotating) plus voltage balancer		rotating rectifier or other brushless machine plus stationary frequency changer and wave shaping

Note: Driven machine for each of the three systems will fit in space presently available.

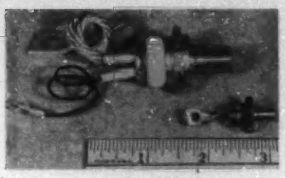


Fig. 2. Silicon p-n-p-n switches: larger one rated 50 amp, 300 volts maximum; smaller one rated 16 amp, 400 volts maximum

There are still other schemes in this category which can successfully pass through synchronous speed and can operate either above or below or at synchronous speed. Such schemes are called double range. The slip channel of these schemes must be able to handle power in either direction, so they are inherently capable of magnetizing the system from the rotor side. The Scherbius system of Fig. 1(B) falls in this group.

In all the nonwasteful synchronous flux schemes there must be at least one commutator-type machine, or its equivalent, which handles power at least as great as the slip times the power output of the system.

When full-rated systems, shown in Fig. 1(C), are considered, there is no question of operation above or below synchronous speed, and usually there will be no need to reverse the direction of power flow in the commutator. However, if the commutator is not capable of 2-way power flow, then it is not possible to supply reactive loads.

Of the schemes so far discussed, the first, Fig. 1(A), has been eliminated because it is wasteful. The two remaining, Figs. 1(B) and 1(C), both use a-c commutators which have to handle a substantial amount of power. Such commutators, as is well known, are not workable in aircraft-type environments; before an aircraft vscf system becomes practical, some other means of performing the commutating function must be used.

Commutating Devices

Table II lists a number of different devices which can be used to perform the commutating function. In the first column, three components are listed as being 2-way conducting and controllable. This means they are capable of carrying current in either direction and that the instant at which they switch or commutate can be controlled. The rotating machine with commutator is, of course, well known. However, it is eliminated from use in vscf systems because even at sea level its output frequency is limited to well below 100 cps. The mechanical rectifier is a mechanically controlled switch which has seen considerable development in the electrochemical industry. This, too, is unsuitable for use in the airplane environments at the frequencies required in vscf systems. The third device included under this listing is a symmetrical transistor. This device is suitable for operation in aircraft environments, but unfortunately its power rating is so small that it, too, is eliminated for use in aircraft vscf systems.

The remaining components listed in Table II are inherently rectifying devices, so some circuit arrangement using two of the components back to back must be used in order to handle current both into and out of the power source or load.

Included under the heading of uncontrolled rectifier (diode) are the mercury-arc rectifier and the semiconductor rectifier. These components have found a very large application in conversion of a-c into d-c energy. These components suffer from the disadvantage that the instant at which they commutate is controlled entirely by the power circuit voltages. Because of this, the uncontrolled rectifier when used alone cannot produce anything other than d-c energy and because of this cannot result in a vscf system.

Fig. 4. Simulated controlled rectifier with rack for 18 units and cooling blower

Under the heading of controlled rectifier are included such devices as the grid-controlled mercury-arc rectifier, the Ignitron, the Thyratron, and such new semiconductor devices as the transistor and the p-n-p-n switch.[5,6] With any of these components, the instant of firing can be delayed by any amount. This means both controlled rectifier operation and inverter operation are possible. That is, power can be handled in either direction by circuits using these components. As has already been seen, capability of handling power in either direction results in a vscf system which can be double range and which can be magnetized from the low-frequency side.

Intermediate between the uncontrolled and the controlled rectifier is an uncontrolled (diode) rectifier together with some form of modulator. Such a combination is capable of a limited amount of delay. Because of the limit on the amount of delay controlled rectifier operation is possible, but not inverter operation. That is, power can flow in only one direction. This, as we have seen, leads to a vscf system which is single sided and which must

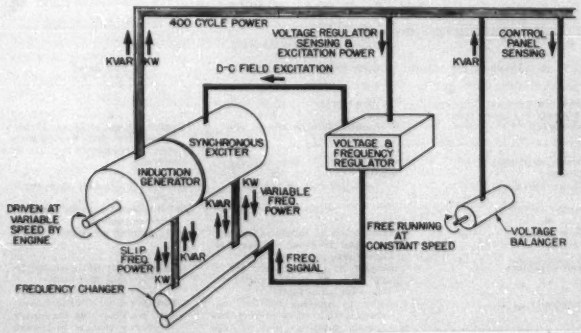


Fig. 3. A vscf system block diagram

Fig. 5. Wound-rotor induction generator and synchronous exciter on the drive stand. Both machines are NEMA 284 frame size

Fig. 6. Frequency changer with control and test instrumentation

be magnetized from the high-frequency 400-cycle side.

The foregoing analysis leads to the conclusions that if there are no commutating devices of any kind in the power handling circuits of the system, then the scheme is a wasteful one.[7] If there is a commutator machine in the power handling circuits, then either the frequency or the speed range capability of the system is severely limited. If uncontrolled rectifiers together with some kind of modulator are used in the power-handling circuits, then there must be capacitors to magnetize the system, and the scheme can be either a full-rated or single range synchronous-flux system. If controlled rectifiers are used, then the scheme may be a full-rated system or either single or double range synchronous-flux system, and capacitors are not needed for magnetization.

Comparison of VSCF Systems

A comparison of nonwasteful vscf systems is shown in Table III.

Three systems are compared: a full-rated system; a synchronous-flux system which operates above and below synchronous speed; and a synchronous-flux system which is confined to one side of synchronous speed.

From considerations illustrated in Table III and because of the power ratings of presently available semiconductor components, it is concluded that the best system for application as an aircraft main generator is a double-range synchronous-flux system. As has previously been pointed out, such a system must use controlled rectifiers, and the only available controlled rectifier which is suitable for aircraft environments is the silicon p-n-p-n switch. This device is now available in ratings large enough to meet most present aircraft requirements as shown in Fig. 2. In fact, it is the advent of this component which has made aircraft vscf systems practical at this time.

Description of the System

A block diagram of the system is shown in Fig. 3.

There are two machines on the variable speed shaft: one the synchronous-flux wound-rotor induction generator; the other, a salient-pole synchronous exciter which in combination with the semiconductor frequency changer handles the slip power. The salient-pole synchronous exciter is an inside-out machine with the d-c field winding on the stator and the a-c output winding on the rotor. The frequency changer is located inside the shaft and rotates with it. This results in a brushless configuration similar to the well-known rotating-rectifier alternator.

A 400-cps reference-frequency signal is provided by a tuning fork or oscillator housed in the same package as the voltage regulator. This signal controls the frequency changer.

The block diagram shows one block called a voltage balancer. This is a free-running synchronous machine floating across the 400-cps bus bars. It is in reality a synchronously rotating amortisseur, whose prime function is to maintain balanced voltages when the load currents are unbalanced. In addition, it supplies some reactive kva to the bus, helps in shaping the 400-cps output voltage wave, and contributes to the fault capability of the system.

Description of the Hardware

Studies along the lines already indicated showed that a vscf aircraft main generator system would be practical and competitive in size and weight, when silicon p-n-p-n switches became available in power ratings of several kw.

Instead of waiting for these components

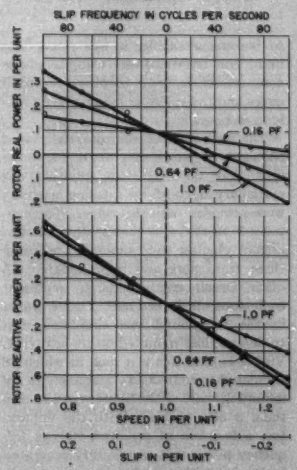


Fig. 7. Frequency changer power output

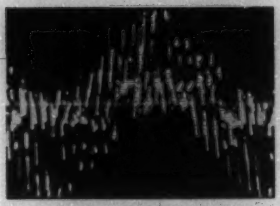


Fig. 8. Oscillogram of the slip frequency voltage fabricated by the frequency changer

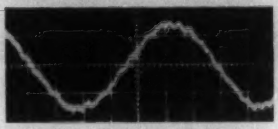


Fig. 9. Oscillogram of the slip frequency current flowing in the frequency changer

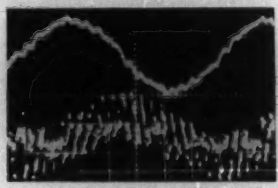


Fig. 10. Oscillogram of slip frequency current and voltage showing current leading the voltage by approximately 100 degrees

to become available, it was decided to build and test hardware using available components, in order to prove the principles and to explore as many of the problems as possible.

The first system was essentially that shown in the block diagram of Fig. 3 except that the frequency changer was stationary and both the rotating machines were provided with slip rings. The controlled rectifiers required for the frequency changer were simulated by using three transistors and approximately 15 other circuit components to represent each controlled rectifier. Fig. 4 shows in the foreground one simulated controlled rectifier mounted on its heat sink. Visible behind this is a rack in which 18 simulated controlled rectifiers are mounted; part of the blower used to cool them is also visible. With this apparatus, a few kva of constant 400-cps frequency power was successfully generated from a variable speed shaft.

As soon as it was determined that the available silicon p-n-p-n switches were sufficiently advanced in their development, a second set of hardware was built, using this component in the frequency changer, and incorporating all the improvements resulting from work with the first set of hardware. Fig. 5 shows the two rotating machines mounted on the drive stand; these are standard commercial machines and are provided with slip rings. Fig. 6 shows the stationary frequency changer together with the control and test instrumentation.

The system shown in these two illustrations, Figs. 5 and 6, is essentially that shown in the system block diagram of Fig. 3, except that voltage regulation is performed manually, the frequency changer is stationary, and the two rotating machines are provided with slip rings.

Test Results

An extensive program of tests was carried out using the hardware shown in Figs. 5 and 6. From these tests the following conclusions can be drawn.

1. A double-range synchronous-flux vscf generator system for a speed range of 2:1 or more is feasible, and in particular that the difficulties of passing through synchronous speed have been overcome.

2. The frequency of the power generated follows that of the frequency-reference signal exactly, no matter what the speed of rotation of the shaft may be, with frequency transients which amount to no more than wave distortion, and phase shift. This means that the integrated frequency error of the system is the same as that of the reference, which can be made very small, and that the frequency transient is so much reduced that a new definition is required.

3. The system behaves substantially

Fig. 11. Oscillogram showing 400-cps output current and voltage at 0.65 power factor

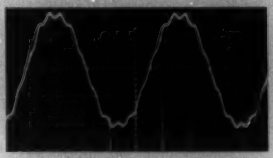


Fig. 12. Oscillogram showing output voltage with d-c excitation of induction generator rotor

in the manner predicted. Fig. 7 is a plot of measurements of the real and reactive power output of the frequency changer for various 400-cps loads and power factors, over a speed range. These plots agree, within the limits of experimental error, with theoretical predictions.[8]

4. The frequency changer fabricates the correct slip frequency voltage and current and can handle real and reactive power flow in either direction. The measurements plotted in Fig. 7 show the latter. The oscillogram of Fig. 8 shows the slip frequency voltage that is fabricated by the frequency changer, Fig. 9 shows the slip frequency current, and Fig. 10 shows the current and voltage on the same oscillogram with the correct

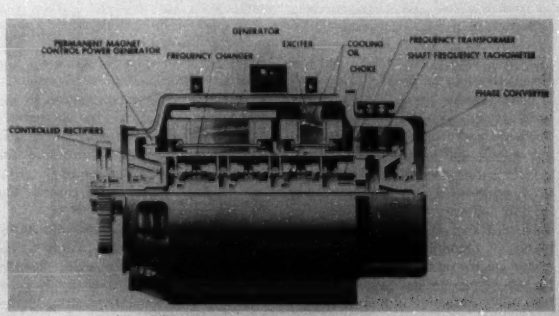


Fig. 13. A 60-kva vscf generator, oil-cooled

phase relationship, the current leading the voltage by about 100 degrees in the case shown.

5. Acceptable 400-cps output voltage wave shape results. Fig. 11 is an oscillogram of the 400-cps output current and voltage with a lagging load of about 0.65 power factor, taken without the voltage balancer connected to the systems. The current wave shape is quite acceptable, the voltage wave shape contains some ripple; however, a major part of this ripple is due to slot harmonic voltage in the induction generator and could, therefore, be eliminated by proper design of this machine. Fig. 12 is an oscillogram of the 400-cps output voltage of the induction generator when its rotor is excited with direct current, where the slot harmonic voltage is clearly visible.

Actual System Designs

As a result of the extensive test program with the laboratory system, it has been possible to make extrapolations to actual aircraft systems with a fair degree of accuracy. A number of complete aircraft systems have been designed. Fig. 13 is a cutaway view of a 60-kva 4,500/8,000-rpm oil-cooled machine.

Conclusion

From designs made, it is concluded that practical vscf generator systems of this type can be built using the silicon p-n-p-n switches which are now becoming available. These systems are competitive in size and total weight with present hydromechanical constant-speed drive and synchronous generator combinations, and they have better efficiency, much better frequency regulation, and will likely prove to be more reliable than the present constant-speed drive and generator combination.

References

1. General Specification for Air-Cooled A-C, Aircraft Generators and Regulators. *Military Specification, MIL-G-6099A (ASG)*, Department of the Air Force and Bureau of Aeronautics, U. S. Navy, Mar. 25, 1957; Amendment 1, Apr. 15, 1958.

2. Polyphase Commutator Machines (book), B. Adkins, W. J. Gibbs. Cambridge University Press, Cambridge, England, 1951.

3. The Nature of Polyphase Induction Machines (book), P. L. Alger. John Wiley & Sons, Inc., New York, N.Y., 1951.

4. Synchronous-Flux Generator, O. J. M. Smith. *Electrical Engineering*, vol. 77, 1958, pp. 605–10.

5. A Silicon-Controlled Rectifier, D. K. Bisson, R. F. Dyer. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, May 1959, pp. 102–06.

6. High-Current Transistors, F. S. Stein, E. W. Torok. *AIEE CP58-1397* (available on request).

7. The Limitations of Induction Generators in Constant-Frequency Aircraft Systems, E. Erdelyi, E. E. Kolsterowicz, W. R. Miller. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, Nov. 1958, pp. 348–52.

8. Energy Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems, M. Riaz. *Ibid.*, vol. 78, Mar. 1959, pp. 25–30.

Discussion

J. T. Duane (General Electric Company, Erie, Pa.): The authors are to be complimented on an interesting and informative discussion of their approach to the generation of constant frequency power with variable speed machines. Generator configurations of this type are receiving increasing amounts of attention throughout the industry and the experimental results reported here provide a useful supplement to some of the theoretical work already published.

Previously available discussions of vscf systems using induction generators, as mentioned in references 7 and 8 of the paper, have presented careful and complete analyses of power flow and distribution within the induction generator. Unfortunately, equal attention has not been given to the problems associated with the generation and transfer of reactive power. Since the authors are reporting experimental data, their comments on the problems associated with reactive power distribution should be extremely interesting.

It is necessary to supply significant amounts of lagging reactive power to induction generators whenever they are operated at slip other than zero. In typical induction motors this reactive power requirement represents a significant percentage of total rating. Do the authors feel that this is the case in the systems they describe, and can they offer an estimate of the reactive power requirements for the induction generator as a percentage of system output?

All reactive power in the system shown in Fig. 3 must be supplied by either the synchronous exciter or the voltage balancer, at other than zero slip. It would appear desirable to minimize that portion of the reactive power generated by the synchronous exciter, since those kilovars (kvar) must be transferred through the frequency changer and the induction generator thus increasing the size of both of these elements of the system. Have the authors reached any conclusions as to what a desirable balance is for the division of reactive power generating capability between the synchronous exciter and the voltage balancer?

The paper points out that the use of a floating synchronous machine on the line can provide amortisseur action to yield balanced voltages under unbalanced load conditions. This will require that the floating machine have a low value of negative sequence reactance. The authors suggest that this can be obtained with a relatively small machine. Experience in the design of conventional synchronous alternators suggests that ohmic values of negative sequence reactance are closely coupled to machine size. Do the authors feel that special or unusual design techniques will be required to build a voltage balancer with a negative sequence reactance as low as a conventional generator without the device being almost as heavy as the conventional generator, and are such techniques already available?

If the voltage balancing function can be provided by a very small machine, will it have the power-dissipating capability necessary to permit extended periods of operation under conditions of unbalanced system loads?

Bert V. Hoard (Boeing Airplane Company, Seattle, Wash.): The authors have written a very good paper on a complex subject. However, there are some areas on operation and performance of the system which, if amplified, will improve its value. The authors state that the difficulties of passing through synchronous speed have been overcome. Have the tests shown any tendency toward roughness, or harmonics of slip frequency in output voltage, or current at or near synchronous speed, as the generator rotor slips or commutates from one phase to another in a slip cycle?

The oscillogram trace of slip frequency voltage, slip frequency not stated, in Fig. 8, as fabricated by the frequency changer, indicates a considerable amount of certain harmonics although the current flowing (Fig. 9) has a materially lower percentage of harmonics. How do these harmonics affect the weight and efficiency of the frequency changer and the exciter at, say, minimum speed and at synchronous speed?

The frequency changer real and reactive power output curves, Fig. 7, are quite interesting. At minimum speed, 25% slip, and unity power factor load on the generator, it appears that more reactive power must be supplied by the frequency changer to the generator rotor than real power. And when the load power factor is lagging at 64%, the reactive power required is almost 100% greater than the real power requirements at unity power factory. Is it correct to conclude that reactive requirements of the generator rotor have more influence in determining the rating and weight of the exciter and frequency changer than real power requirements of the generator rotor both at the minimum speed?

Normal loading for the semiconductors in the frequency changer that is considerably below their rating, probably has been necessary to prevent failure during abnormal generator loading conditions. These abnormal conditions may be 200% load, or balanced or unbalanced faults, followed by removal of the abnormal load and then transient high over voltage occurs. Have any approximate derating factors been determined for the semiconductors to prevent failure during these abnormal conditions?

What is the expected efficiency of the 60-kva generator, Fig. 13, when operating at rated load and power factor at minimum speed and at synchronous speed?

K. M. Chirgwin and **L. J. Stratton**: Mr. Duane raises the problem of the flow and distribution of reactive power in the induction generator. This problem is analyzed in reference 8 of the paper, but from the questions that have been asked it seems that this problem is not well understood.

The behavior of an induction machine

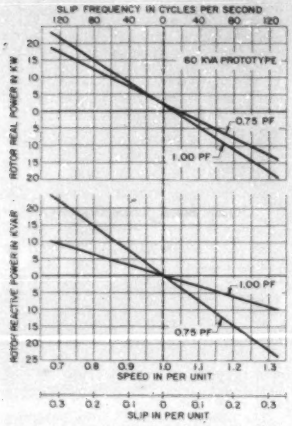

SLIP FREQUENCY IN CYCLES PER SECOND
120 80 40 0 40 80 120
ROTOR REAL POWER IN KW
60 KVA PROTOTYPE
0.75 PF
1.00 PF
ROTOR REACTIVE POWER IN KVAR
1.00 PF
0.75 PF
0.7 0.8 0.9 1.0 1.1 1.2 1.3
SPEED IN PER UNIT
0.3 0.2 0.1 0 0.1 0.2 0.3
SLIP IN PER UNIT

Fig. 14. Rotor power requirements of the induction generator

when its rotor circuits are excited by an external source differs from its behavior when its rotor is short-circuited or closed through resistances. In particular the excited induction machine is capable of generating reactive kva. A fraction s of the kvar generated in the stator must be supplied to the rotor from the external exciting source, but the balance or a fraction $1-s$ times the kvar generated originates inside the induction machine itself, where s is the per-unit slip. Operation at synchronous speed is not an isolated point but merely the point at which s is zero and where all the kvar originates inside the induction machine. Above synchronous speed more kvar than is needed is generated inside the induction machine and the extra kvar is absorbed by the external exciting source.

The curves of Fig. 14 have been prepared to show the real and reactive power requirements of a 60-kva flyable prototype system which our company is presently building.

Mr. Duane asks what is a desirable balance between generating kvar in the voltage balancer on the stator side and generating kvar in the synchronous exciter on the rotor side of the induction generator. The voltage balancer is a constant speed, probably high-speed, machine, probably operating in a favorable environment and so can undoubtedly put out a given kvar for less weight than can the synchronous exciter. On the rotor side of the induction generator, however, only a fraction s of the kvar is needed, so this favors the exciter, but handling these kvar increases the size and weight of the frequency changer and the induction generator. So we prefer to make the trade off between the voltage balancer and the main machine package, which comprises exciter, frequency changer, and induction generator.

In making such a trade off all the following functions of the voltage balancer should be considered, since if they are not performed by the voltage balancer then the weight of the main machine or other components of the system must be increased:

1. Voltage balance
2. Supply of kvar
3. Output wave shaping
4. Fault kva capability

If the trade off is based only on considerations of kvar then it can be said that low synchronous speed of the voltage balancer, now a synchronous condenser, high minimum speed of the main machine, and narrow speed range of the main machine, all tend to reduce the proportion of the kvar that should be supplied by the voltage balancer. In the limit this will lead to a constant speed synchronous alternator where, as is well known, no synchronous condenser is indicated.

In vscf systems designed to meet military specifications the size of the voltage balancer is determined by the voltage unbalance requirements. The data given below refer to the voltage balancer which forms part of the 60-kva flyable prototype system which is presently being built. The nominal rating of the voltage balancer is 15 kvar. On the 15-kvar base, the negative sequence quantities are: reactance 7.45%, resistance 4.15%, impedance 8.45%.

The electromagnetic weight of this machine, while it is greater than that of a conventional 15-kva synchronous condenser, is only about one half the weight of a conventional 60-kva alternator for the same speed and environment. The design to achieve this is unusual because the requirements are unusual, but the design techniques are no different from those used for normal synchronous machines. The thermal capability of the amortisseur of this voltage balancer is such that it can carry continuously a single-phase current equal to one third of the normal full load current of the system.

Mr. Hoard asks if there is any tendency toward roughness in the output voltage when operating very near synchronous speed.

The operation of the frequency changer is the same whether the frequency being fabricated is 100 cps or a fraction of a cps so synchronous speed does not represent a discontinuous point and there is no tendency toward roughness in passing through this speed. During demonstration of our vscf hardware we have shown that, with the drive stand speed set as close as possible to synchronous speed, the output wave form of the frequency changer automatically compensates for speed irregularities caused by the drive stand governor. In fact, the frequency changer fabricates the most precise wave when the ratio of input to output frequencies is the greatest. Therefore, operation near synchronous speed represents the best condition since the ratio is approaching infinity.

As can be seen from Fig. 8, the fabricated slip frequency voltage which is approximately 70 cps, contains harmonics, with 1,400 cps being the largest component. These harmonics result from the fabrication process wherein a low-frequency voltage is fabricated from a high-frequency voltage by switching in pieces of the high-frequency waves.

A choke is included in each phase of the frequency changer to absorb the harmonics in the same way as a choke is used to filter the output of a d-c rectifier power supply. The effectiveness of these chokes is illustrated by the smoothness of the slip frequency current wave. The weight of the chokes is determined by the minimum speed since the harmonic frequencies are the lowest for that condition. From other considerations, the weight of the exciter is also determined by the minimum speed and the additional harmonic power represents only a slight increase in the total exciter kva, and only a slight increase in exciter weight.

The harmonics do not materially affect the frequency changer efficiency since the forward voltage drop of the controlled rectifiers is in the order of a volt and the slight increase in current contributed by the harmonics does not represent many watts. Because the controlled rectifiers are either "full-on" and are "naturally commutated" to "full-off," they provide the most efficient switching process. It is important that any semiconductor frequency changer be operated in this manner since if wave shaping is accomplished by absorbing voltage across the switching element, appreciable power is dissipated in the switch with attendant loss in efficiency and with considerable derating of the switching element. Therefore, it is better to allow harmonics to be produced in the conversion or frequency changing process and to filter them out, rather than avoid producing them in the conversion process by causing high power dissipation in the switching elements.

The ratio of real and reactive power values shown in Fig. 7 presents a somewhat distorted picture since the wound rotor induction generator used for the data was a commercial machine having a magnetizing reactive power requirement equal to 1.6 times its output kva rating. More typical requirements for a 60-kva system with a 15-kvar voltage balancer are shown by Fig. 14. The rating and weight of the frequency changer is determined by the maximum kva which occurs at either the minimum or maximum speed.

The amount of derating of the semiconductors in the frequency changer depends not only on meeting overload and fault conditions but also on cooling air velocity and temperature. Presently available controlled rectifiers are rated for a continuous peak inverse voltage of 300 volts and a current of 50 amperes which corresponds to 15 kw. Since 18 controlled rectifiers are used in the frequency changer, and only approximately 30 kva need be converted by the frequency changer for a 60-kva machine, adequate switching capacity is available in presently available controlled rectifiers.

Calculated efficiencies for a 60-kva vscf air cooled machine similar to that shown in Fig. 13 including voltage balancer losses are: 82.2% at minimum speed and 81.6% at synchronous speed for rated kva and unity power factor load; and 77.1% at minimum speed and 77.4% at synchronous speed for rated kva and 0.8 power factor load.

A High-Capacity Maintenance-Free Generating System for Motor Coaches

R. L. LARSON
ASSOCIATE MEMBER AIEE

THE TREND of electric loads on mobile equipment has been upward at an ever-increasing rate since the first generators were installed on automobiles in 1909.

In the past the load-trend curve has roughly followed that of the electrical utility industry which doubles in load every 10 years. But the adoption of air-conditioning equipment in recent years has increased the connected load on automotive-type equipment far beyond the value which would have been predicted by the load-trend curve. The highest electric load on commercial vehicles is encountered on city coaches. The growth of the electric load for coaches with 12-volt systems is shown in Fig. 1.

The capacity of the generator used to supply the load increased from 50 amperes in 1933 to 120 amperes in 1946. However, even this striking increase in output does not indicate the relative capacities of the two generators since the 120-ampere unit began supplying current to the load at 550 rpm, while the 50-ampere unit required a speed of 900 rpm to produce useful output. The speed at which the generator begins to supply current to the system became more important with the adoption of the hydraulic transmission, in as much as the normal pattern of operation in large cities is to operate the engine at full throttle for a period measured in seconds, then coast to the next stop.

Until this year, the standard generating system for most large city coaches consisted of a 120-ampere d-c generator and double voltage and current regulators of the vibrating contact type. The generator was gear driven at a 1.94 ratio and produced 80 amperes at engine idle speed. The field circuit was split into halves with each half regulated by a separate voltage and current regulator, so that twice the normal field strength could be used without exceeding the maximum current rating of the regulator contacts.

Several things happened that determined that the d-c generator system was headed for obsolescence on large city coaches. In addition to the rising load trend curve heading toward a point where the electric capacity and speed range of the largest conventional d-c generators would be inadequate, the diesel engines were being improved to the extent that engine overhauls were made less and less frequently. No longer would the generator and regulator operate satisfactorily without maintenance over the lengthened engine overhaul periods. In fact, the increased electric loads and the high engine compartment temperatures, which resulted from increased engine loads, contributed to the reduction in life of the brushes, oil seals, bearings, and other components from that which previously had been obtained. Alternating-current generators with selenium rectifiers had been applied on some coaches with higher than normal electric loads. However, although these generators would solve the electric capacity problem for several years, the existing units had the same mechanical construction deficiencies as the d-c generator and, in addition, required the use of a large, expensive, and vulnerable selenium rectifier.

Design Objectives

About 1955, the engineers at General Motors Truck and Coach Division began setting down specifications for the type of generating system which would be needed in the future. Because this generating system was expected to be one which would supply the needs of the coach operators for a number of years without any major revisions, the requirements set down were not met by any generating system in existence. A goal of 1 million miles of maintenance-free operation was set up to correspond to the expected life of the engine. The minimum generator capacity was set at 200 amperes, with 125 amperes output available at engine idle speed with provision to increase this output in the future with only minor revisions in the generator. The generator was to be of the a-c type with built-in silicon rectifiers and, in order to obtain the long life operation desired, would have to be brushless and have permanently lubricated bearings. To prevent failures caused by the entry of dirt, detergents, and other foreign matter into the generator or failures caused by lack of proper maintenance on air filters designed to prevent such damage, the generator was required to be environment free. The regulator was required to be of a completely static design to eliminate all wearing parts and was to have the ability to retain its voltage setting unchanged over a period of years.

In order that maximum performance could be obtained from a given size generator, the regulator was also required to supply much higher field power than had previously been thought practical.

Paper 59-779, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing April 10, 1959.

R. L. LARSON is with the Delco-Remy Division, General Motors Corporation, Anderson, Ind.

The author wishes to acknowledge the contribution of the following people who were instrumental in the development of this system: the original conception of such a system and much of the early work on the generator was done by R. H. Bertsche, R. O. Ellerby, E. B. Gegenheimer, and R. A. Carlson, of the General Motors Truck and Coach Division. The transistor regulator development was done by L. R. Hetzler and L. J. Sheldrake and the refinements in design and later development work on both the generator and the regulator carried out by L. J. Raver; all three of these men are with the Delco-Remy Division of General Motors.

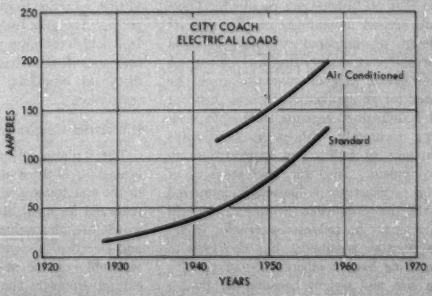


Fig. 1. Load-trend curve of motor coaches

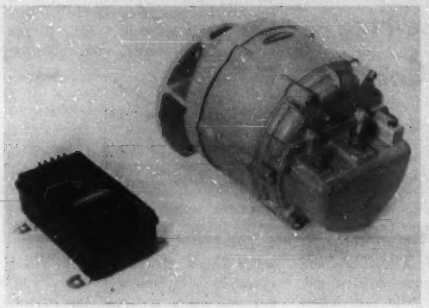


Fig. 2. Transistor regulator and oil-cooled generator

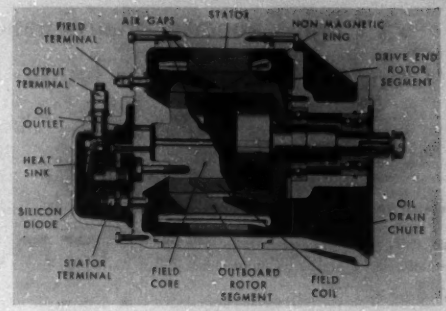


Fig. 3. Cutaway view of the oil-cooled generator

Generator Design

The generator and regulator designed to meet the objectives specified is shown in Fig. 2, and a cutaway view of the generator is shown in Fig. 3. Basically, the machine is a 3-phase alternator with a 12-pole Lundell-type rotor. To obtain the brushless feature, however, several modifications have been made in the normal Lundell rotor construction. Two bearings are used on the drive end so that the entire rotor is supported from that end in an overhung fashion. A nonmagnetic ring has been added between the six poles of the drive-end rotor segment and the six poles of the outboard rotor segment, so that the outboard segment is supported mechanically from the drive-end segment although the two are still isolated magnetically. An interior air gap has been added to separate the field coil and core assembly from the rotor poles, and the field coil and core assembly is supported from the end frame opposite the drive end.

In this manner, a rotating field machine is made in which both the field winding and the output winding are stationary so that no brushes are required. There are several advantages of this type of design as compared with other types of brushless generators that are known. A permanent magnet generator could be used for an application such as this, but it is difficult to regulate the output voltage accurately without a considerable amount of bulky and expensive extra equipment. Also the field strength of such a generator is limited as compared with that obtainable in an excited field machine. An inductor alternator would be another possibility, but this type of machine has an inherent 2-to-1 penalty in size and weight because the flux is pulsated from a maximum to a minimum value instead of being alternated between positive maximum and negative maximum values. A rotating rectifier type of machine could also be used but, in the size of generator considered here, an exciter is not needed to amplify the excitation power available from the regulator to a level sufficient to excite the main field. Consequently, the exciter becomes merely added weight, space, and expense which serves no purpose other than to transfer the excitation power from the stationary member to the rotating member.

An additional advantage of the double air-gap Lundell construction as compared to the normal Lundell construction arises from the separation of the rotating member from the field coil and core assembly. This reduces the moment of inertia of the rotor from 125 pound-inches^2 for a comparable size generator of standard Lundell construction to 96 pound-inches.^{2} This reduction in inertia is important since the generator-to-engine gear ratio is limited to a value which will not cause shaft and gear damage from the torsional vibrations of the diesel engine. Because of this reduction in inertia, it was possible to raise the gear ratio from 1.94 to 2.29. It would have been possible to raise the ratio even further had not the gear center distances available imposed a further restriction.

Rectifier Construction

The 3-phase full-wave bridge rectifier is made up of three silicon diodes mounted in the end frame and three silicon diodes mounted in small aluminum heat sinks, which are themselves supported from, and electrically common with, the three terminals of the stator winding which extend through the end frame. The aluminum heat sinks also act as junction blocks for connecting the leads of the three end-frame-mounted diodes to the stator terminals. All six of the diodes are of the same polarity in this construction.

Oil Cooling System

Several attempts were made to make the generator environment free by utilizing totally enclosed or totally enclosed fan-cooled construction but it was found that at the ambient temperatures prevailing in the engine compartment, such a machine became too large and too heavy for consideration. Water cooling was also considered, but this involved the use of tubes and passages to conduct the water to and from the heat sources in the generator and was thought to be too expensive and too vulnerable to leaks to be considered. The only other cooling medium available on the engine was the engine oil. Since the oil is not electrically conductive, it can be thrown directly on the windings without fear of causing electrical failure, and it has the added advantage of furnishing a permanent lubrication supply for the bearings. For these reasons, oil cooling was adopted for this generator.

The rectifier cover casting has provision to receive a fitting from an oil line which is connected to the engine oil gallery. There is a small orifice drilled in this casting which restricts the flow of oil to approximately one gallon per minute at the 40 pounds per square inch regulated oil pressure which is available when the engine is operating in its normal speed range. The orifice also acts to produce a high velocity jet of oil down one side of the rectifier chamber, which keeps the oil in the chamber in motion and prevents large temperature differences around the

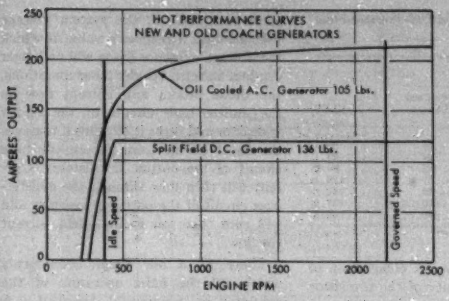


Fig. 4. Comparison of generator output curves

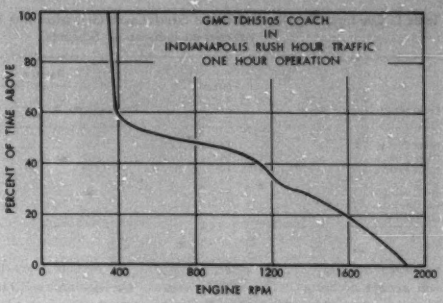


Fig. 5. Engine operating speed characteristic

diodes. With the diodes entirely submerged in oil, excellent heat transfer is obtained so that large heat sinks are unnecessary. The oil fills the rectifier chamber to the level of the oil outlet. It then flows through the oil outlet and through the hole in the center of the field core. At the end of the field core the oil is discharged against the inner surface of the rotor and is carried outward by centrifugal force to the inner air gap, through which it flows toward the rectifier end of the machine. Some of the oil escapes through slots between the nonmagnetic ring and the drive-end rotor segment and follows the underside of the poles outward until it is thrown off the poles onto the inside of the stator winding at the drive end. The rest of the oil continues toward the rectifier end of the machine until some of it escapes through the slots between the nonmagnetic ring and the outboard rotor segment. The remainder continues through the air gap to the end of the rotor where it is thrown on the inside of the stator winding at the rectifier end. Thus, the oil cools the stator winding by direct contact with the end windings and cools the field winding by conduction of heat through the core assembly and by direct contact with the field winding at the inner air gap. It then drains to the bottom of the stator assembly and, in the case of the oil at the rectifier end, flows through a notch between the stator laminations and the stator frame and drains down the chute in the drive-end frame to the flywheel housing of the engine. The oil level in the generator remains below the height of the air gap so that the rotor does not pick up the oil which has drained to the bottom of the stator.

The oil is taken from the engine after it has passed through the full-flow filter and the oil cooler in order that the cleanest and the coolest oil in the system will be used in the generator. Even under these conditions, however, the oil can at times reach an inlet temperature of 240 F (degrees Farenheit), although the most normal operating range is around 200 F. The temperature of the oil increases about 50 F in passing through the generator under full load conditions at high speed. One of the major problems to be solved in the development of the oil cooling system was the selection of insulation which would withstand the high temperature oil bath. The insulation system being used consists of an epoxy coating on the magnet wire with glass tape on the field coil and isocyanate-impregnated glass slot insulation in the stator. All assemblies receive two coats of an oil modified phenolic varnish which was selected after extensive testing of all available varnishes in a 300 F bath of used engine oil. The components used in the insulation system all have successfully withstood more than 12,000 hours in this bath. No difficulty has been experienced in operating units from insulation failures due to deterioration of the insulation.

When the oil cooling system was first proposed, it was thought that the generator would act as a centrifugal and magnetic oil filter and would soon be rendered inoperative by an accommodation of sludge. These fears have proved groundless, even when the experimental generators were operated on old engines with bypass oil filters that permitted quite dirty oil to flow through the generator. The heads of the bolts which extend into the rectifier chamber from the field core are highly magnetized and attract any iron particles which come through the oil system. The only particles discovered on the bolt heads have been particles of approximately 1 micron in diameter which pass freely through the oil filter. Particles of this small size should not cause damage even if they pass completely through the generator.

Performance

A comparison between the performance of the new a-c generator system and the d-c generator system which it replaced is shown in Fig. 4. This illustrates the increase in output of approximately 75% throughout the speed range and the 23% reduction in weight which was obtained. Part of this increase in performance, of course, is the result of the 18% increase in gear ratio on the new system.

The importance of obtaining high generator output at engine idle speed becomes apparent if the curve of engine operating speed versus percentage of time shown in Fig. 5 is examined. This particular engine speed curve was taken on a coach operating in downtown Indianapolis, Ind., during a normal afternoon rush hour, but it is representative of the conditions encountered in most medium size cities. Surveys were made of the operating conditions in metropolitan New York City and New Orleans, La., during the rush hour a few years ago which indicated that the engine operated at idle speed as much as 70% of the time.

The engine operating speed characteristic can be combined with the generator performance curve, the generator drive ratio, and the battery characteristics to determine the amount of load which can be carried by the equipment in question on that particular driving schedule without producing a net charge or discharge of the battery. This value of system load is referred to as the "city load-carrying capacity" (CLCC). In determining the CLCC, an assumption must be made as to the state of charge and the temperature of the battery since this de-

Table I. City Load-Carrying Capacity of Transit Coach Generators with Two 205-Ampere-Hour Batteries on Indianapolis Schedule

Generator	Gear Ratio	Voltage Regulator Setting	Battery		CLCC
			Temperature, F	Per-Cent Charge	
Split Field	1.9375	14.5	−10	60	88.7
Split Field	1.9375	14.5	−10	80	81.1
Split Field	1.9375	14.5	+100	80	98.7
Oil Cooled	2.29	14.0	−10	80	124.1
Oil Cooled	2.29	14.0	+100	80	168.5
Oil Cooled	2.29	14.0	+100	100	137.9
Oil Cooled	1.9375	14.0	+100	80	142.0

termines the rate at which the battery will accept a charge. The battery temperature also determines the efficiency of the battery in ampere-hours of output as compared to the ampere-hours of input. The CLCC rating of a generator gives a much more accurate measure of its load carrying ability under variable-speed driving conditions than does the maximum output rating. A comparison of the CLCC rating of the two coach generator systems with various battery conditions is given in Table I. The split-field generator calculations were made on the basis of a 14.5-volt regulator setting because that is a normal setting. The oil-cooled generator calculations are based upon a 14.0-volt setting because the battery can be kept charged with the lower setting when the higher capacity generator is used. The lower system voltage lengthens the life of lights and other electric equipment and reduces battery maintenance.

Regulator Design

The only regulating unit that is necessary with this oil-cooled generator is a voltage regulator. No current regulator is required since the generator is inherently current limited and no cutout relay is required since the silicon diodes effectively block any reverse current. The elimination of these two components from the regulator eliminates the need for heavy leads and large terminals to handle the output current and consequently reduces the size and weight of the regulator greatly. A comparison of the size and complexity of the transistor regulator used with the oil-cooled generator and the split field regulator used with the d-c generator and the carbon-pile regulator used with the older style a-c generators is shown in Fig. 6. The comparison in size is even more striking when it is considered that the transistor regulator will handle two to three times as much field power and give much closer voltage regulation.

Circuit

The circuit of the transistor regulator is shown in Fig. 7 and the physical arrangement of the components on the printed circuit board is shown in Fig. 8. The voltage divider network at the left side merely provides a means of adjusting the regulator voltage setting and of compensating for variations in the Zener voltage of the voltage reference diode. When the system voltage is higher than the regulator setting, the voltage reference diode will pass current, and this current flows in the emitter-base circuit of the driver transistor, turning it "on." In the "on" condition, the emitter-collector circuit of the driver transistor is essentially a short circuit which is connected across the emitter-base circuit of the two output transistors. This effectively short-circuits any base current which would tend to flow in the output transistors and turns them "off." Since the output transistors are in series with the generator field winding, no field current can flow. With the field current turned off, the system voltage will fall until it reaches a value at which the voltage reference diode will no longer conduct current. Under those conditions, the diode blocks any current flow in the emitter-base circuit of the driver transistor and turns it off, thus it removes the short circuit from the emitter-base circuit of the output transistors. Current will then flow through the emitter-base circuit of the output transistors and will turn them on so that field current can flow.

There are a few circuit refinements which aid the basic operation of the regulator as previously described. A silicon diode is connected between the positive line and the emitters of the output transistors so that field current flows through it in the forward direction. This is referred to as the back-bias diode. A high-resistance back-bias bleed resistor is connected from the negative side of the diode to the negative line. The purpose of this combination is to prevent excessive leakage through the output transistors when they are in the "off" condition. The driver transistor is not a perfect short circuit in the "on" condition and would tend to allow a small amount of base current to flow in the output transistors, which would in turn allow a high leakage current to flow through the field. The drop across the driver transistor is approximately 0.3 volt in magnitude. By bleeding a small amount of current through the back-bias diode to develop its threshold voltage of approximately 0.5 volt, the voltage relationship at the output transistors is established with the bases being slightly more positive than the emitters. This biases the emitter junctions in the reverse direction and prevents any base current from flowing, thereby reducing the leakage current to a very low value. The nonlinear forward characteristic of the back-bias diode is used in place of the linear characteristic of a resistor to avoid dissipating a large amount of power in the regulator and reducing the power available for excitation of the field.

The feedback resistor is a low-value resistor which will have a voltage drop of approximately 0.1 volt across it when full field current is flowing. Its function is to cause the switching action of the regulator to occur very rapidly so that the output transistors are either in the "full on" or "full off" condition. It accomplishes this by subtracting a small increment of voltage from the voltage reference diode when the field current is turned on and adding a small increment of voltage to the voltage ref-

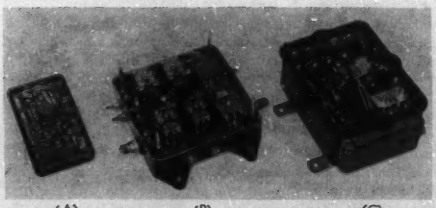


Fig. 6. Regulators used on motor coaches
A—Transistor voltage regulator
B—Split field voltage and current regulator with cutout relay
C—Carbon pile voltage regulator with cutout relay

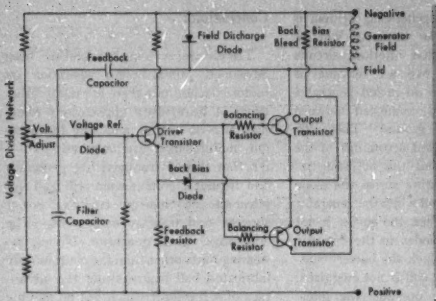


Fig. 7. Schematic circuit diagram of the transistor voltage regulator

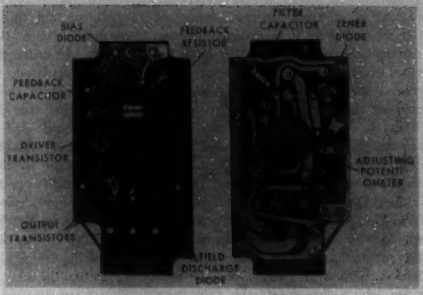


Fig. 8. View of both sides of the printed circuit board assembly from the transistor voltage regulator

erence diode when the field current is turned off. This mode of operation is necessary so that the output transistors will go rapidly through the "partially on" condition, which is the condition of maximum heat dissipation. If the transistors were to operate by modulating the field current rather than by switching it, either the heat sinks provided would have to be of an excessively large size or the current rating of the unit would have to be reduced drastically. The feedback capacitor aids in securing this sharp switching action by transferring a voltage pulse from the output circuit back to the voltage sensing circuit.

The field-discharge diode is connected so that it is in parallel with the generator field coil. When the generator field is turned off suddenly, a high voltage will be induced in the field winding if no path is provided to dissipate the energy stored in the magnetic field. This high voltage would puncture the output transistors. The field discharge diode provides a low impedance path through which the field current can circulate to discharge the stored energy but does not dissipate any power when the field current is on since its leakage current is quite small. The time constant of the field winding with this low external resistance in series is high enough that the regulator does not switch in an extremely rapid manner as would be the case if a discharge resistor were used. Since the absence of light flicker in the coach is obtained with this regulator by the very small value of voltage excursion rather than by a high rate of switching, this low switching rate is desirable to limit the transistor heating.

The filter capacitor in the voltage-sensing network has the function of limiting the ripple which appears at the voltage reference diode. The action of the regulator is so fast that it would be actuated by each ripple peak on the voltage and would regulate at the peak value of the ripple instead of at the average value. This is not desirable since the magnitude of the ripple changes with load and frequency so that controlling the ripple peaks would result in variations of the average voltage of the system. In automotive application, there is nearly always a battery on the system which acts as an excellent ripple filter, and thus, aids this filter capacitor in performing its function.

The switching action of the regulator is illustrated by the oscillograms of field voltage, field current, and terminal voltage in Fig. 9. It is readily apparent how the relative time duration of the "on" and "off" periods varies with the load on the system. Because of this action, the regulator is referred to as being of the pulse time controlled type. The temperature characteristics of the voltage reference diode and of the driver transistor are approximately equal and opposite so that the regulated voltage does not vary appreciably with changes in ambient temperature. This is a desirable characteristic for a coach regulator. Because the regulator is mounted at a considerable distance from the battery, its temperature does not reflect battery temperature, for which many automotive regulators are compensated.

The ripple which appears on the field voltage and terminal voltage waves is 60 cycles, picked up by the oscillograph. When this ripple is eliminated, the terminal voltage excursions during switching are found to be about ±0.1 volt and the variation in the average voltage with changes in load and generator speed does not exceed 0.4 volt over a temperature range of −20 F to +160 F. This represents much better voltage regulation than has been available on any automotive-type equipment in the past.

Coach Circuit

The generator and regulator described are markedly different from those which have been used in the past and consequently the application of these units to the coach brings up some problems that have not existed previously.

A d-c generator and a cutout relay are normally applied with the shunt field connected to the generator side of the cutout relay. When the engine is started, the generator builds up enough voltage to close the cutout relay and begins to supply current to the system. When the

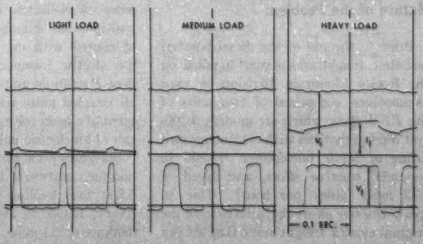


Fig. 9. Oscillograms of field voltage V_t, field current I_f, and terminal voltage V_t of oil-cooled generator with transistor regulator in operation

engine is stopped, current from the battery will flow in a reverse direction through the cutout relay to drive the generator as a motor. This reverse current is used to open the relay and disconnect the generator from the system. With the a-c generator and the silicon rectifier, however, a cutout relay is not used. Since the field must be connected on the d-c side of the rectifier, there is nothing to prevent the battery from furnishing field current even when the engine is not in operation. In order to prevent discharge of the batteries by this field current, a field relay must be provided which disconnects the field from the battery when the engine is shut down. This relay is normally actuated from the ignition switch of a gasoline engine or from the "run" switch of the diesel engine.

The voltage appearing across the cutout relay contacts is frequently used to operate various interlock circuits on the coach, in as much as no voltage can appear at that point when the engine is operating and the generating system is normal because the contacts will be closed. In order that interlock circuits can be actuated, one a-c terminal is brought out of the oil-cooled generator and a relay coil is connected between this terminal and ground. Thus, the relay coil is connected across one of the silicon diodes so that the pulsating reverse voltage appearing across the diode actuates the relay coil when the generator is in operation. When the engine is not in operation, the diodes in the opposite arm of the bridge block the battery voltage so that the relay coil is not energized. This relay in turn operates the interlock circuits through normally open or normally closed contacts as desired. This connection was tried previously with selenium rectifiers and was unsuccessful because the high leakage of the selenium cells would sometimes allow the relay to remain actuated even though the engine was shut down, but the very low leakage of the silicon cells makes this system practical.

Conclusions

The electric system described has been developed primarily for operation on motor coaches, but there are many other types of heavy-duty applications where high electric output and long life without maintenance are required. It is felt that this environment free generator and regulator combination will find application on marine engines, power shovels, and military equipment. The advantage of a generator having no wearing parts other than the continuously lubricated ball bearings and the advantage of a regulator which has no moving parts to wear out or change adjustment have been recognized by all operating personnel having contact with this system. Its reception has been gratifying.

Experimental units have been on test since December 1956 and limited production has been under way since February 1958. The results of these operating units indicate that this approach is basically sound.

Multiple-Unit Operation of Diesel and Electric Locomotives on the Milwaukee Road

LAURENCE WYLIE
FELLOW AIEE

THE USE of a diesel-electric locomotive booster unit in conjunction with one or more straight electric locomotive units was conceived to meet specific requirements of heavy freight train operation on the electrified Rocky Mountain Division of the Milwaukee Road; see track profile, Fig. 1.

Nature of the Problem

Prior to the use of the diesel booster, fast-time freight trains were handled on the Rocky Mountain Division by road locomotives, comprised of two units of the *EF-4* class, which are modern 3,300-volt d-c locomotives with continuous unit rating of 5,110 horsepower, 78,300 lb (pounds) tractive effort, and speed of 25.2 mph (miles per hour). Due to speed–tractive effort characteristics of the original electric freight units, they are not suitable for this service and, even if they were suitable, they could not be operated with more than four units in multiple and could not be operated in conjunction with the modern *EF-4* units due to the differences in control system and speed characteristics.

The problem under discussion is that of synchronizing the operation of one or more diesel-electric units equipped with multiple-unit engine-load-regulator type of control with one or more 3,300-volt d-c electric locomotives equipped with type *P* multiple-unit control, and to have all coupled units respond to the movement of a main controller in the operating cab of an electric unit.

Six years prior to the use of the diesel booster, eastward freight trains with 5,800 tons trailing were handled from Avery, Idaho, to Deer Lodge, Mont., a distance of 211 miles, by road locomotives consisting of two *EF-4* units coupled as the road locomotive. A helper locomotive, consisting of four *EF-1* units was, and still is, used on the 1.7% grade between Avery, Idaho, and Haugan, Mont., a distance of 38 miles, and over the Continental Divide between Butte Yard and Piedmont, on 1.6 and 2.0% grades, a distance of 36 miles. Prior to the use of the diesel booster unit, it was necessary to reduce a train of 5,800 tons to a train of 4,800 tons at Deer Lodge in order to eliminate the use of a helper locomotive on the long 1.0% grade between Lombard and Loweth, a distance of 48 miles. The use of the diesel booster eliminated necessity for train reduction at Deer Lodge and improved the running time, even with 5,800 tons, between Deer Lodge and Harlowton.

Due to the demand for increased speed of westward time freight trains with 3,600 tons trailing, it was necessary to increase the tractive effort and horsepower of the road locomotives used on these trains to such an extent that a speed of

Paper 59-492, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Chicago, Ill., April 7–9, 1959, and presented at the Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted December 15, 1958; made available for printing January 30, 1959.

LAURENCE WYLIE is a Consulting Electrical Engineer, Seattle, Wash.

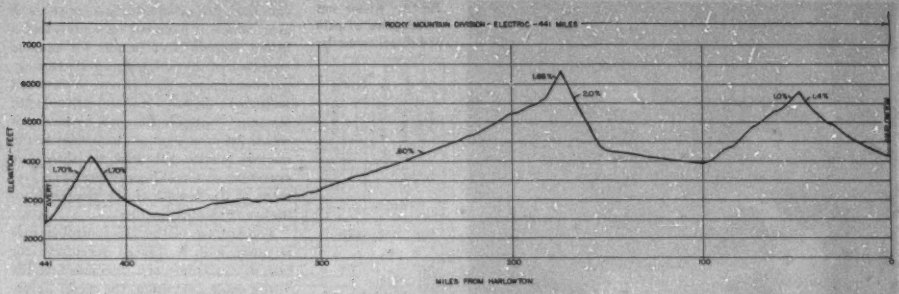


Fig. 1. Profile of Rocky Mountain Division

approximately 25 mph could be made while ascending the 2.0% grade between Piedmont and Donald, a distance of 21 miles, without the use of a helper locomotive, and with consequent increase in train speed at other locations where track condition and curvature would permit.

As the principal advantage of the use of the diesel booster occurs between Deer Lodge and Harlowton, with either eastward or westward trains, the booster unit is occasionally operated only between these two points.

Equipment for Control of Booster Unit

The control of the diesel booster unit is accomplished in the following manner:

1. A standard 27-wire control cable is installed in each unit of the electric locomotive and control circuit completed between the two electric units and the diesel unit by means of standard diesel receptacles and jumper cables.

2. A miniature diesel controller (hereafter referred to as the diesel synchronizing, or DS controller) is mounted on the forward side of each main controller in the operating, or uncoupled, ends of the electric units. These miniature controllers are connected by a rack and pinion to the main controller handle of the electric units in such manner that movement of the electric controller handle will rotate the drum of the DS controller in proper sequence.

3. Microswitches, to control the operation of the reverser on the diesel unit, are mounted inside the case of the main controller on the electric locomotive.

4. A miniature diesel meter and switch panel is installed at the right side of the electric locomotive meter panel. This diesel panel contains the following equipment:

Ammeter, indicating diesel traction motor current.

Switches for control of equipment in diesel unit as follows:

Control and fuel pump
Engine run and controller
Generator field
Headlight
Sanders
Engine stop button
Alarm bell
Wheel slip light

DS Controller Mounting and Mechanical Connection to Main Electric Controller

The DS controller is mounted on a heavy sheet iron bracket which is secured firmly to the main electric controller case by heavy cap screws at the forward end of the controller braking segment; see Fig 2. A composite view of the main electric controller, diesel controller, and meter panels, is shown in Fig. 3.

Rotation for positioning of the DS controller cam shaft is accomplished by means of a gear rack connected to the main electric controller handle by means of a heavy "Z" bar bolted to the controller handle and a roller bearing rod end on the rack bar. This rack bar is held in mesh with the small gear on the diesel controller by a spring-mounted roller bearing on top of the diesel controller case. The back of the rack bar is milled to a predetermined profile which is governed by movement of the electric controller handle and eccentric mounting of the gear on the DS controller. The eccentric mounting of the gear and the irregular movement of the rack bar combine to rotate the shaft of the diesel controller through 222 degrees with a corresponding movement of only 42 degrees of the electric controller handle while moving from *OFF* to the 16th notch, or series parallel position, and through only 108 degrees while the electric controller is moved from the 16th

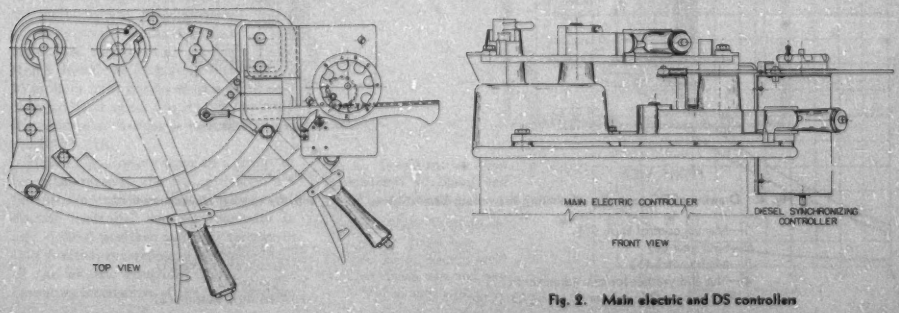


Fig. 2. Main electric and DS controllers

Fig. 3. Meter and switch panel of main and diesel controller

notch to the 37th notch, or an additional 45 degrees, to the full parallel position.

Description of DS Controller

A front and side view drawing of the DS controller, with cover removed, is shown in Fig. 4. The controller consists of a welded-sheet-metal box 6 inches wide, 5 inches deep and 8 inches high, containing a roller-bearing-mounted shaft on which is mounted six Dilecto fiber disks 1/4 inch thick and $3^1/_2$ inches in diameter. Five of these disks are cam-cut to control the operation of microswitches mounted on a removable rack, and the sixth disk serves as a stop and positioning device to prevent undesired rotation of the cam shaft.

On the protruding upper end of the cam shaft are mounted a small spur tooth gear, which is mounted off-center, and a heavy Dilecto fiber handle for manual operation of the switch in case of necessity or emergency. This handle is fastened securely to the cam shaft and is fitted with a manual control locking pin, by means of which the spur gear can be locked to the manual handle and shaft for automatic operation of the DS controller by movement of the handle of the electric locomotive controller handle.

The bottom of the DS controller case is fitted with a 6-point receptacle for ease in connecting and disconnecting the 6-wire cable containing the diesel governor control wires.

Electrical Characteristics of Diesel Control Circuits and DS Controller

A schematic diagram of the diesel synchronizing control circuits is shown in Fig. 5. Control current for operation of the diesel unit is supplied by the battery on the diesel unit. The microswitch in the circuit controlling the diesel generator field contactor must rupture a current of 0.80 ampere and, for this

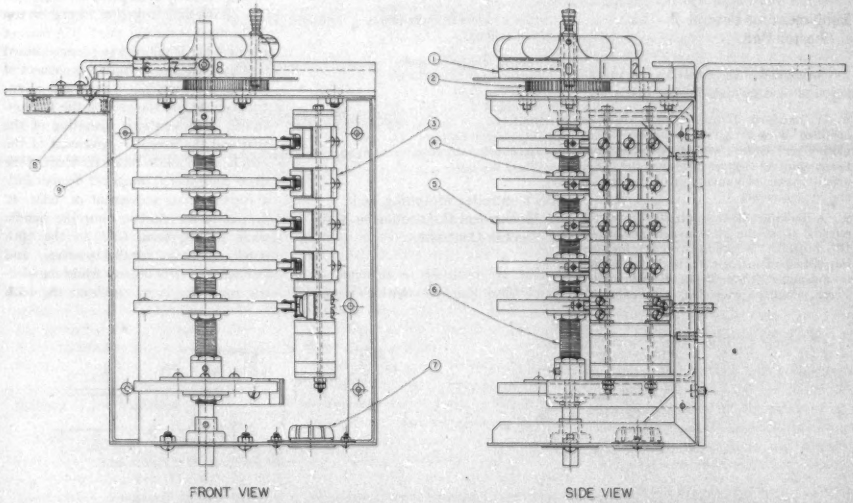


Fig. 4. Drawing of DS controller illustrating the various parts as given below, with the number of each required

1—Manual control knob (1)
2—Spur gear (1)
3—Microswitch (5)
4—Nut and washer for cam adjustment (12)
5—Linen base Dilecto cam 1/4x$3^1/_2$ O. D. (6).
6—Cam shaft ((1)
7—Amphenol socket (1)
8—Bearing (1)
9—Cam shaft bearing (2)

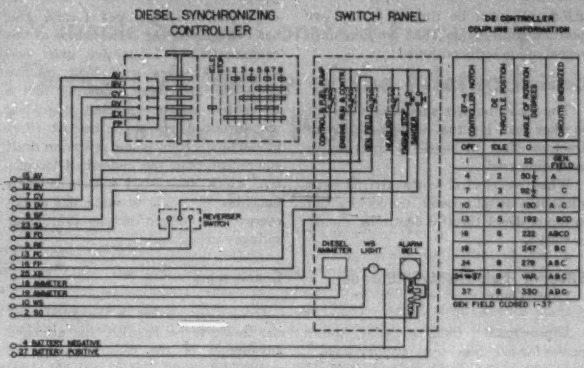


Fig. 5. Schematic diagram of DS control circuits

reason, this switch is provided with a magnetic blowout. The other four microswitches each carry a current of only 0.10 ampere and so do not require magnetic blowout protection. All microswitches have roller bearings on the operating lever. The outer ends of the operating levers require a movement of less than 1/8 inch, and the cams are cut accordingly to a depth of 1/8 inch.

The DS controller is designed to supersede the functions of the controller in the diesel unit when the diesel unit is being used as a booster with one or more electric units. The load on the diesel unit is regulated by energizing a single coil, or various combinations of coils, in the load governor on the diesel unit. In order to avoid overloading the traction motors and wheel slipping on the diesel unit, the Run 8 throttle position is not reached until the electric controller handle has been advanced to the 18th notch, or just beyond the series parallel position, and corresponding to a speed of approximately 15 mph, or well above the point where wheel slipping might normally occur on the diesel unit. The diesel unit then continues to operate in Run 8 throttle position at all speeds above 15 mph, or until the diesel unit is cut out, changed to manual control, or the electric controller handle is retarded below the 18th notch.

In starting a heavy train on a grade, or for other reasons, the DS controller can be disconnected mechanically from the electric controller and set manually in any desired position and will remain in that position regardless of the movement of the electric controller handle. The operating mechanism of the DS controller can manually be locked to the electric controller mechanism at the will of the operator, and can only be so locked in the proper position.

Regeneration and Dynamic Braking

Extra wires are provided in the 27-wire control cable to provide for control of dynamic braking on the diesel unit if this should be necessary or advisable. At the present time, dynamic braking is not used on the diesel booster unit as the electric road and helper locomotives have sufficient capacity to handle the train while descending the heavy mountain grades and, in this way, secure credit for the electric energy which is returned to the power supply system or absorbed by trains in the vicinity which are using power.

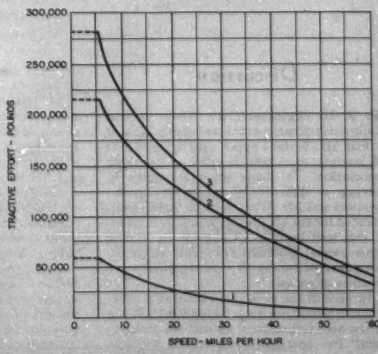


Fig. 6. Speed-tractive effort characteristics of diesel and electric units

1—One EMD 1,750-horsepower diesel unit
2—Two EF-4 electric units
3—One diesel unit and two EF-4 electric units

Preparation and Operation of Multiple-Unit Diesel and Electric Locomotive

The diesel locomotive should be coupled to the back end of a type *EF-4* electric locomotive after which the procedure should be as follows:

Air

Couple brake pipe and signal line hoses and open hose cocks. Couple equalizing pipe hose on diesel to independent application pipe hose on electric locomotive and open hose cocks. Do not couple main reservoir equalizing pipe hose or sanding pipe hose. Condition brake equipment on diesel for "trailing" movement.

Control Jumpers

Install diesel-type control jumper between diesel and electric locomotive, and also between electric locomotives if two electric locomotives are used.

Control Transfer

Proceed as follows:

In electric locomotive operating cab, on diesel control panel in cowl in front of of engineer, close the following switches:

1. Control and fuel pump.
2. Engine run.

In diesel locomotive operating cab, on engineer's panel, open all switches except automatic sander. In diesel control cabinet, all switches must be in ON position and isolation switch in RUN. Headlight controller switch, located below window in front of engineer, must be placed in 9 o'clock position, marked "Controlled from another unit coupled at either end." Control of diesel locomotive is now transferred to operating cab of electric locomotive.

CUTTING IN DIESEL LOCOMOTIVE

Close engine run and generator field switches, located in diesel control panel in cowl in front of engineer in electric locomotive operating cab. Diesel locomotive will then follow the operation of the electric locomotive controller and provide tractive effort according to position of the controller.

CUTTING OUT DIESEL LOCOMOTIVE

Open engine run and generator field switches, located in diesel control panel in cowl in front of engineer in electric locomotive operating cab.

NORMAL OPERATION: MOTORING

Throw the reverser on *EF-4* controller for desired direction of movement and then advance the controller handle in the usual manner, allowing sufficient time for slack action in the train and for throttle governor control in the diesel unit. The diesel throttle controller will follow the movement of the *EF-4* controller so that the diesel locomotive will be in no. 8 throttle, or full run position, when the *EF-4* controller is partially out. During further advance of the *EF-4* controller, the diesel locomotive continues in no. 8 throttle position. Any running position on the *EF-4* controller is a running position on the diesel locomotive.

If the diesel locomotive begins slipping for any reason, it will automatically sand and ultimately unload itself. Before this condition develops, action should be taken to correct by partially unloading the diesel locomotive.

REGENERATION

Before commencing regeneration, the *EF-4* controller handle must be moved to the OFF position and engine run and generator field switches on diesel panel in the *EF-4* cab must be thrown to the OFF position and left in that position until regeneration is discontinued.

GENERAL

The diesel booster should only be cut in when there is actual need for additional tractive effort to handle a train on an ascending grade.

When an electric locomotive is used without a diesel booster, the locking pin in top of diesel throttle controller should be raised.

Effect of High Tractive Effort on Draft Gear

Approximate tractive effort–speed curves for one diesel unit, two *EF-4* electric units, and for three units combined, are shown in Fig. 6. Starting tractive effort at 25% coefficient of adhesion for the three units is 277,000 pounds. At 25 mph, the combined output at the rail is 11,500 horsepower.

Mechanical characteristics of couplers applied to locomotives and freight cars are as follows:

1. Older type *E* couplers with *E 50* knuckles, grade *B* steel normalized average 225,000 lb:
   Yield point, 560,000 lb.
   Ultimate strength, 600,000 lb.
   Safe load, 200,000 lb.

2. New type *E* couplers with *E 50* knuckles, normalized and tempered:
   Yield point, 300,000 lb.
   Ultimate strength, 535,000 to 630,000 lb.
   Safe load, 250,000 lb.

3. Type *E* couplers for diesel or electric locomotives with type *E* quenched and tempered high-tensile-strength knuckles:
   Yield point, 700,000 to 750,000 lb.
   Safe load, 350,000 lb.

The use of the two *EF-4* units with the diesel booster has not caused failures of draft gear on locomotives or cars. Our observations on the performance of this combination of locomotive units and other locomotives with high tractive effort and horsepower, indicates that a locomotive which can exert a gradual and sustained increase in tractive effort in starting a train is less severe on draft equipment than the undersized locomotive which must start a train by first bunching the slack to make use of the energy thus stored in the draft gear springs and cushions.

Conclusions

Due to the nature of the profile and the traffic to be handled, the principal advantages of the diesel booster are realized between Deer Lodge and Harlowton, on both eastward and westward trains. On this subdivision, the increase in gross ton-miles per train-hour which can be attributed to the use of the diesel booster with full-tonnage trains is 13,759 for eastward trains and 25,700 for westward trains. The more favorable showing on westward trains is due to the fact that on these trains there is a much greater percentage in the increase in train tonnage, with the prevailing grade descending, as compared to ascending for eastward trains.

The performance of the auxiliary diesel control equipment has been satisfactory. The use of a single diesel booster unit with two electric units comprising a road locomotive for through service, with helper locomotives on heavy grades as required, has improved freight train performance by combining the characteristics of the two different types of motive power, and by providing additional tractive effort where required for heavy grade operation.

Discussion

W. S. H. Hamilton (New York Central Railroad (retired); now Larchmont, N. Y.): I read Mr. Wylie's paper with a great deal of interest, especially in view of our long association. It shows great ingenuity in solving a difficult if somewhat special problem and Mr. Wylie is to be highly commended for the solution at which he arrived. It can well be used as a guide in solving other special problems that arise which are not provided for by the standardized locomotives of today.

In view of the problems which were overcome, it would not surprise me if Mr. Wylie in the near future came up with an arrangement for operating locomotives in two different locations in the train using wireless for control.

As one of the original instructors of enginemen at the time of the electrification of the Milwaukee Road in 1915, I used to know nearly every foot of the right-of-way of the section shown in Fig. 1. The weight of train handled as given in the paper has greatly increased since those days, the original trailing tonnage being 2,500, as I remember it. Also, train speeds have just about doubled, which is to be expected in order to keep up with modern-day progress.

I was glad to note that the freight car couplers have also kept up with progress to a degree that permits concentrating the amount of drawbar pull, that the locomotives described can exert, at the head end of the train. In my day helper locomotives had to be cut into the middle of the train, and even then one had to be careful in starting lest some drawbar, or "lung" as the trainmen called them, came out by the "roots."

Laurence Wylie: Mr. Hamilton's comments are appreciated greatly, especially in view of his long and varied experience with electrical problems associated with railroad transportation. It is regretted that more written discussion of this paper was not submitted by others.

The matter of remote control of electric or diesel locomotives placed at various locations throughout long trains is receiving attention by this author, as well as by others now interested in such proposed operation.

Variable-Speed Constant-Frequency Devices: A Survey of the Methods in Use and Proposed

T. B. OWEN
MEMBER AIEE

EVER since the advent of a-c systems in aircraft, designers have been increasingly led to consider methods of obtaining constant frequency from variable-speed sources, as the aircraft engine is usually at least two to one in speed range. The first device to obtain constant speed from a variable-speed source was under development for many years, and it is only recently that it has been considered reliable enough for general use. Fortunately, the timing has been good, since the very large increase in aircraft electrical loads has coincided almost exactly with the availability of this device. Of course, the large increase in electrical loads has led to the proposal of many devices either to provide constant speed to drive a generator, or to obtain constant frequency directly from a generator driven at a variable speed.

The Mechanical Differential

Before surveying the methods proposed, the mechanical differential will be examined in some detail. It will become apparent as this discussion progresses that the differential method outlined is employed, in one form or another, in most of the proposed devices. If, for example, a constant output shaft speed of 6,000 rpm is required, and a variable input speed of from 4,000 to 5,000 rpm is available, it is clear that some type of variable-ratio gearing must be interposed between the input and output shafts. One possible means of doing this is shown in Fig. 1, where the speed ratio between the input and output shafts depends upon the position of the roller upon the cone. This is a very common method, and is frequently seen in applications such as the step pulley on lathes.

Another possibility is that shown in Fig. 2, the planetary gear system. Here, where there are two input shafts, the speed of the output shaft depends upon the direction of rotation and the speed of the two input shafts. The differential shaft may, of course, be turned by any of several methods. One method is to run a variable-displacement hydraulic pump from the input shaft. A fixed displacement hydraulic motor then drives the differential shaft at the speed fixed by the setting of the variable-displacement pump. The output shaft speed will be the sum of the input speeds. On the other hand, if the input shaft speed is above the desired output shaft speed, the motor and pump exchange functions. The output shaft speed is now the difference between the two input shaft speeds. It is important to see here that in the first instance the differential shaft *adds* power as well as rpm, while in the second case, the differential shaft *subtracts* power as well as rpm. An electric generator-motor combination, of course, may be used in lieu of the hydraulic-pump–motor combination.

A third possibility is shown in Fig. 3. Here, the variable-speed input shaft drives a variable-displacement hydraulic pump to which is fixed the body of a fixed-displacement hydraulic motor. The motor output drives the output shaft directly. The output shaft speed is the sum of the input speed and the speed of the motor relative to that of the pump. This is, of course, fixed by the wobble-plate setting of the pump. Again, where the input speed is higher than the output shaft speed, the functions of pump and motor are reversed, and the output shaft speed is the difference between the input speed and the motor-pump speed relative to the input. The hydraulic-pump–motor combination could have been replaced here also by an electric generator-motor combination.

Mathematical Relations

It is now necessary to set up the mathematical relationship which relate the input, output, and differential torques, powers, and speeds. For this, the following nomenclature will be used:

w_i = angular input speed
w_o = angular output speed
w_{dir} = angular component of speed transmitted directly, as that of the input shaft of the planetary gear arrangement of Fig. 2
w_{dif} = angular component of speed transmitted through the differential, as in Fig. 2

T_i, T_o, T_{dir}, T_{dif} apply similarly to torque
P_i, P_o, P_{dir}, P_{dif} apply similarly to power
f_i, f_o, f_{dir}, f_{dif} apply similarly to frequency

From an examination of Fig. 2, the planetary gear arrangement, it is apparent that between the input and output there are two speed-torque combinations that add or subtract to equal the output.

1. An input torque T_{dir} which is transmitted at the input speed, or w_{dir}.
2. A differential torque T_{dif} which is transmitted at the differential speed w_{dif}.

Now neglecting efficiencies, and considering only where the input speed is below the output:

$$w_{dir} + w_{dif} = w_o \tag{1}$$

$$w_{dir} T_{dir} + w_{dif} T_{dif} = w_o T_o \tag{2}$$

$$P_{dir} + P_{dif} = P_o \tag{3}$$

Equation 3 is simply another way of

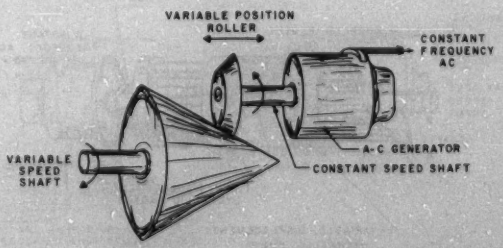


Fig. 1. Variable-ratio gear to provide constant speed from variable speed

Paper 59-780, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 16, 1959; made available for printing April 10, 1959.

T. B. OWEN is with the Douglas Aircraft Company, Santa Monica, Calif.

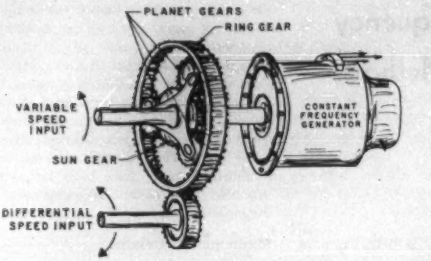


Fig. 2. Planetary gear arrangement for differential speed

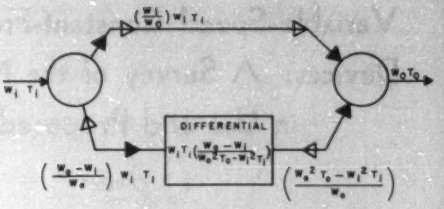


Fig. 4. Schematic diagram of power flow

Black arrow: below synchronous speed
Light arrow: above synchronous speed

stating equation 2, as power is the product of torque and speed.

Further consideration will also show that the direct, differential, and output powers are proportional to the speeds. That is, if $w_{dir} = 3{,}000$, $w_{dif} = 3{,}000$, and $w_o = 6{,}000$, one half the output power is supplied by the differential and one half by the direct path. This relationship may be summarized as follows:

$$w_{dir} : w_{dif} : w_o = P_{dir} : P_{dif} : P_o \qquad (4)$$

Now if the case where the input speed is above the output is considered, the following is obvious:

$$w_{dir} - w_{dif} = w_o \qquad (1A)$$

Then, further consideration will show that the power equation becomes:

$$P_{dir} - P_{dif} = P_o \qquad (2A)$$

That is, both power and speed are subtracted. If the differential speed and power are considered negative where $w_i > w_o$, equations 1 and 2 stand as they are.

To see what the foregoing equations actually mean, consider a case where $w_o = 6{,}000$ and $P_o = 100$. In one case, let the input $w_i = 3{,}000$; in the other, $w_i = 9{,}000$. Since $w_i = w_{dir}$, in each case $w_{dif} = 3{,}000$. Then, making the proper substitutions in equation 4:

Case I

3,000: 3,000: 6,000 = P_{dir}: P_{dif}: 100

$P_{dir} = 50$

$P_{dif} = 50$

Case II

9,000: 3,000: 6,000 = P_{dir}: P_{dif}: 100

$P_{dir} = 150$

$P_{dif} = 50$

The significance of this is that where the input speed is below output speed, the output power is the sum of the direct and differential powers; where the input speed is above the output speed, the output power is the difference between the direct and differential powers. The differential power is thus returned to the shaft. This is why the pump-motor or generator-motor combinations reverse their roles. This must be accomplished or the efficiency would be very low indeed, for otherwise this power is dissipated in the machine.

The two cases illustrated, where the input speed is either below or above the output speed, have been reduced to a single servo-type diagram which is shown in Fig. 4. All quantities are in terms of the input and output speeds and torques and the direct and differential paths are shown and are labeled as such. The direction and quantity of the power flow are shown by the arrows and labels. The quantity in the differential box is the K or multiplication factor necessary to transform the power in the differential loop to that required.

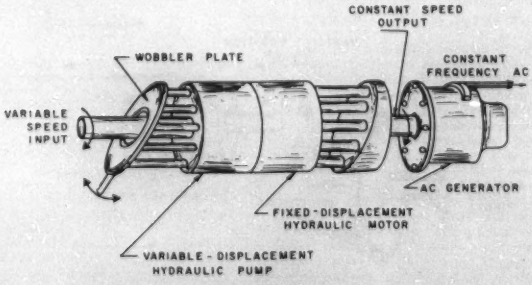


Fig. 3. Hydraulic-motor-pump differential for constant speed

Classification of Methods of Obtaining Constant Frequency from Variable-Speed Sources

In general, three methods are currently in use to obtain constant frequency. First, using some mechanical device, obtain a constant speed and drive a conventional generator. Second, drive the generator at the variable speed and by special methods obtain constant-frequency output. Last, use one of the frequency changers. As will be shown, the first two methods are differential, while the third is nondifferential. For the sake of convenience, the following classification is presented:

Differential

A. Constant-speed output devices:

1. Differential action by means of variable roller dimensions or by variable roller position.

2. Differential action by means of a hydraulic-pump–motor combination.

3. Differential action by means of an electric generator-motor combination.

B. Constant-frequency output devices:

1. Differential action by supplying a-c generator field with slip-frequency power.
   (*a*) Exciter-frequency changer.
   (*b*) Exciter-rotary amplifier.

Nondifferential Methods

A. Frequency changers:

1. Rotary motor generators.

2. Static frequency changers.

CONSTANT-SPEED DEVICES

1. The first class listed includes the mechanical speed changers, several of which are being actively promoted at this time. They

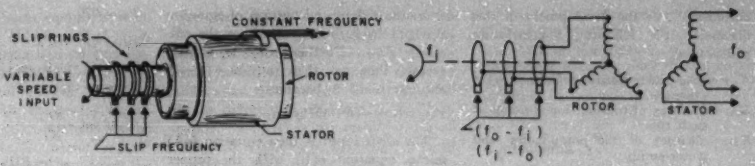


Fig. 5. Constant frequency from variable speed by slip-frequency excitation of rotor

are adequately represented by the device pictured in Fig. 1.

2. The second class listed includes the hydromechanical devices in rather common use today, and is represented in Figs. 2 and 3. Some devices of this type use piston-type hydraulic pumps and motors; others use balls, pistons, or a combination of the two.

3. The main difference in the third type of constant speed device is the substitution of an electric generator-motor combination in place of the hydraulic-pump–motor combination. With this in mind, Figs. 2 and 3 are adequate.

Constant-Frequency Devices

In contrast with the constant-speed devices, these devices use no mechanical auxiliaries. A variable-speed shaft drives the a-c generator, and constant frequency is taken from the generator output. The principle is illustrated in Fig. 5. Here the input shaft speed is w_i generating the frequency f_i. A frequency $(f_o - f_i)$ is fed into the rotor and the output frequency will be the sum, f_o. By proper choice of the frequency fed into the rotor, the desired frequency will appear at the stator output.

It must be appreciated that exactly the same rules apply here as to the mechanical differential. Differential power supplied or extracted is always proportional to the differential speed or frequency which is called the "slip speed" or "slip frequency." The servo-type diagram of Fig. 4 is applicable here as in the case of the mechanical differential. Neglect of this has led many would-be inventors astray.

The exciter-frequency changer is shown in Fig. 6. The exciter and generator rotors are on the same shaft. The exciter frequency is changed from f_i to $(f_o - f_i)$ by the frequency changer and this is fed to the rotor. The output frequency is $f_i + (f_o - f_i) = f_o$. This applies to the case of below synchronism. For above synchronism, f_i is changed to $(f_i - f_o)$ and is subtracted from f_i, $f_i - (f_i - f_o) = f_o$. The frequency changer is not necessarily external to the machine, but may be on the shaft.

The exciter-rotary amplifier is shown in Fig. 7. Here, no frequency changer as such is employed. The field excitation of the exciter is at the differential frequency so that it acts as a rotary amplifier.

Nondifferential Methods

Frequency changers of various types are well known and need no explanation. Several new approaches which utilize solid-state devices seem to show considerable promise. The scheme is either to rectify the variable-frequency alternating current and then produce alternating current of the desired frequency by some variation of the oscillator–power-amplifier principles, or to generate at much higher frequencies than the final desired frequency and then, by switching accurately, to produce the lower frequency.

Development of Efficiency Equations

The first part of this paper encompassed only physical methods of conversion to a constant frequency. In all of the cases noted, no account was taken of the efficiencies, nor was the problem of obtaining excitation power considered. In this section, the efficiencies will be noted and describing equations developed, but no numerical comparisons will be made. The excitation problem will not be dealt with, except superficially. Further, the losses occasioned by auxiliary devices such as pumps or fans for coolant distribution, and friction and windage losses will be neglected entirely. For practical use, therefore, the expressions developed will have to be modified to take excitation, friction, windage, and auxiliary devices into account. Also, electrical and/or mechanical difficulties that would be encountered in making a device operable will not be considered.

Constant-Speed Devices Driving a Generator (Figs. 1–3)

There are two possible modes of operation, below synchronism and above syn-

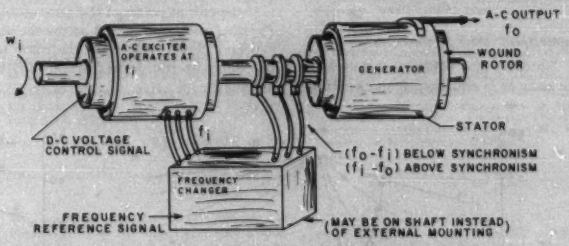


Fig. 6. Differential by frequency changer

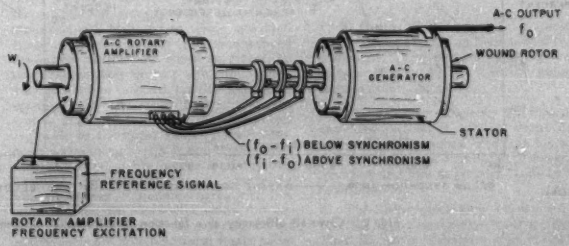


Fig. 7. Differential by rotary amplifier

chronism. For the development of this discussion, the following nomenclature will be used:

K_g = mechanical to electrical efficiency to the generator
K_{dif} = efficiency of transmission of power at shaft rpm
K_{dif} = efficiency of the pump-motor and motor-pump

Input Speed Below Synchronism

If the power out of the drive is P_{do}, the power contributed by the direct and differential portions is as follows:

$$P_{dir} = \frac{w_i}{w_o} P_{do} \tag{5}$$

$$P_{dif} = \frac{(w_o - w_i)}{w_o} P_{do} \tag{6}$$

These two components, with the appropriate efficiencies applied, and with $R = w_i/w_o$, $(1-R) = (w_o - w_i)/w_o$

$$P_{dir} \text{ input} = \frac{RP_{do}}{K_{dir}} \tag{5A}$$

$$P_{dif} \text{ input} = \frac{(1-R)P_{do}}{K_{dif}} \tag{6A}$$

The total drive input is equation 5(A) plus equation 6(A). Then, since the drive output P_{do} is the generator output at the efficiency K_g, the over-all efficiency of this system becomes:

$$\text{Efficiency} = \frac{K_g K_{dir} K_{dif}}{RK_{dif} + (1-R)K_{dir}} \tag{7}$$

Input Speed Above Synchronism

The direct and differential powers are the same as in equations 5 and 6. However, the differential portion of equation 6 is put back into the shaft at the efficiency K_{dif}. Further, since w_i is now greater than w_o, then $(w_i - w_o = (R-1)$ and equation 6(A) becomes:

$$P_{dif} \text{ return} = (R-1)P_{do}K_{dif} \tag{6B}$$

The shaft input is then equation 5(A) minus equation 6(B) with the proper substitution made for P_{do}. The efficiency of the system now becomes:

$$\text{Efficiency} = \frac{K_g K_{dir}}{R + (R-1)K_{dif}K_{dir}} \tag{8}$$

Differential Devices: Constant-Frequency Generator (Fig. 6)

The approach here must be different, as other efficiencies are involved, and the generator is not in series with the constant speed drive. The nomenclature will be

K_g = mechanical to electrical efficiency of the generator
K_{eg} = mechanical to electrical efficiency of the exciter acting as a generator
K_{em} = electrical to mechanical efficiency of the exciter acting as a motor
K_f = conversion efficiency of the frequency converter in either direction
K_t = transformation efficiency of differential power from the rotor to the stator, or vice versa

$w_i/w_o = R$
$\frac{w_o - w_i}{w_i} = (1-R)$ below synchronism
$(R-1)$ above synchronism

Input Speed Below Synchronism

If the output power of the generator is P_o, the power transmitted across the air gap (*ag*) (direct) and from the rotor to the stator by transformer action (*dif*) is as follows:

$$P_{ag} = RP_o \tag{9}$$

$$P_{dif} = (1-R)P_o \tag{10}$$

At the efficiencies involved in each path, these become:

$$P_{ag} \text{ input} = \frac{RP_o}{K_g} \tag{9A}$$

$$P_{dif} \text{ input} = \frac{(1-R)P_o}{K_t K_f K_{eg}} \tag{10A}$$

The total power input is equation 9(A) plus equation 10(A). Adding, and then taking the efficiency:

$$\text{Efficiency} = \frac{K_g K_t K_f K_{eg}}{RK_t K_f K_{eg} + (1-R)K_g} \tag{11}$$

Input Speed Above Synchronism

In this case the differential power subtracted from the armature, $(R-1)P_o$, is put back into the shaft at the efficiencies K_t, K_f, and K_{em}. Equation 10(A) must now be revised as follows:

$$P_{dif} \text{ return} = (R-1)P_o K_t K_f K_{em} \tag{10B}$$

The shaft power input is now equation 9(A) minus 10(B). Making the subtraction, and finding the efficiency:

$$\text{Efficiency} = \frac{K_g}{R + (R-1)K_t K_f K_{em} K_g} \tag{12}$$

Differential Devices: Constant-Frequency Generator (Fig. 7)

If this figure is examined closely, it will be seen that the circuit is very much the same as that of Fig. 6, with the exception that K_f, the frequency converter efficiency, is not present. Thus, if K_f in equations 11 and 12 is taken as unity,

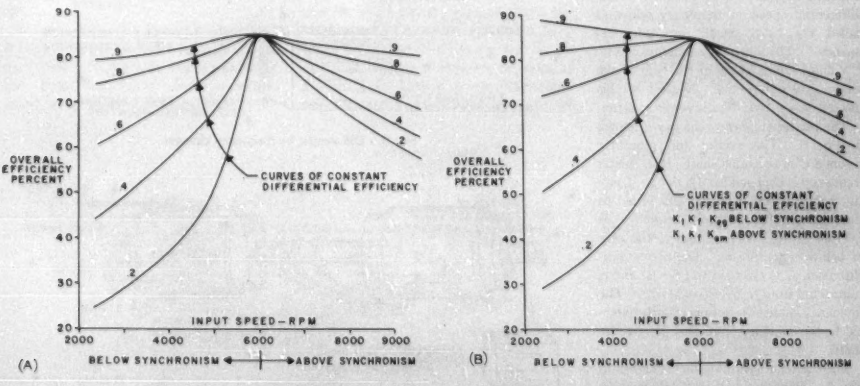


Fig. 8. Over-all efficiency as a function of differential efficiency and input speed

A—Constant-speed device plus generator
B—Variable-speed constant-frequency generator

these equations will serve equally well for the case of Fig. 7. (The fact that K_f is not present in the equations for this system does not necessarily mean that the over-all efficiency will be better. K_{ss} and K_{sm} are different for the two systems, and without direct knowledge of the values in each case, it cannot be said that one would be more efficient than the other on the basis of the equations developed.)

Efficiency Evaluation

With the development of the basic efficiency equations, it is now possible to make an assessment of over-all efficiency as a function of input speed, with certain factors being held constant. These constants will be defined for each of the two cases considered.

Constant-Speed System

Here, constant speed is obtained by a mechanical device and used to turn the generator. Typically, the generator may be assigned an efficiency of 85%. The efficiency K_{slf} may be considered as being unity as it is essentially a gearing efficiency, and is quite high in any event. K_{slf} may be subject to wide variations. With these assumptions, over-all efficiency is plotted in Fig. 8(A) as a function of input rpm for values of K_{slf} from 0.2 to 0.9.

Constant-Frequency System

In this case, no constant-speed device is interposed between the variable-speed source and the generator. Again, generator efficiency is held at 85% over the entire speed range, although it is probable that there would be some variation both above and below this figure. If one examines equations 11 and 12, it is seen that the efficiencies associated with the differential system occur as a group, and if the differential efficiency is held constant, Fig. 8(B) is the result. Over-all efficiency is plotted as a function of input rpm for values of the "differential efficiency" ($K_1K_fK_{ss}$) and ($K_1K_fK_{sm}$) held constant at values from 0.2 to 0.9. Where systems are studied in which K_f does not appear, it may again be treated as being unity, and the curves are still valid.

Comparative Discussion

The notable fact about Figs. 8(A) and 8(B) is that there is such great similarity. However, no account is taken of the actual efficiencies to be expected in the various systems. Therefore, a very brief note of the probable range of each is given in the following.

K_{slf}. This is probably quite high, in the range of 0.8 or above.

K_{sm} and K_{ss}. On the usual synchronous machine with a rotary exciter, the exciter efficiency is commonly in the order of 70% or less because the exciter of such a machine must have a high forcing action to supply the needed excitation for load changes, faults, etc., at the speeds necessary in the modern electric system. A high-efficiency machine does not have the reserve capacity necessary to accomplish this function. Since the exciter in the cases presented here is more than an ordinary exciter, in that it supplies power as well as excitation, it is not clear whether its efficiency will be as low as in the case of the d-c exciter. In any event, it certainly cannot exceed the usual generator efficiency of 85%, and is more likely less than 80% efficient.

K_f. Reliable sources have stated that the efficiency of this device will be 90% or above, which seems reasonable.

K_1. This is essentially a high-reactance transformer efficiency since there is an air-gap between the rotor and the stator. It is difficult to see how its efficiency can exceed 80%.

Conclusions

No definite conclusions can be drawn from the preceding development, and none were intended. The intention was merely to point out that all of the proposed "new" systems are basically very similar to the old ones, and a preference for one over the other will mainly depend upon availability and the state of development. Certainly there is nothing here to indicate radical superiority in any case.

The efficiencies cited here may, in fact, not be realizable on a practical system, as other considerations must be entertained. Further, the most efficient system may be such that it cannot be built to operate both above and below the synchronous speed, and though many such disabilities may occur, it is not the function of this paper to point these out.

References

1. Theory and Calculations of Electrical Apparatus (book), Charles Proteus Steinmetz, McGraw-Hill Book Company, Inc., New York, N. Y., 1917, chaps. 12 and 13.

2. Theory and Calculation of Alternating Current Phenomena (book), Charles Proteus Steinmetz, McGraw-Hill Book Company, Inc., 1916, chap. 19.

3. The Limitations of Induction Generators in Constant-Frequency Aircraft Systems, E. Erdelyi, E. E. Kolatorowicz, W. R. Miller, *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, Nov. 1958, pp. 348–51.

4. Energy-Conversion Properties of Induction Machines in Variable-Speed Constant-Frequency Generating Systems, M. Riaz, *Ibid.*, vol. 78, Mar. 1959, pp. 25–30.

Discussion

Bert V. Hoard (Boeing Airplane Company, Seattle, Wash.): The results given in Mr. Owen's paper are quite interesting. Frequency make-up systems or differential systems, as a class, certainly are not new. The use of low-loss semiconductors with these systems is new and will provide better systems than were previously available in almost every respect. That factor with respect to aircraft electric systems is very desirable.

I have some questions and comments which, I hope, will elicit a clearer definition of the limits of application of the paper and its usefulness to the industry.

1. The efficiencies shown in Fig. 8(B) are for what percentage load and power factor on the generator over the speed ranges shown?

2. Were the exciter efficiencies K_{ss} and K_{sm} assumed constant or variable for the curves of Fig. 8(B)? This is a very important factor since, for some systems of this type, the exciter physical size may equal or exceed that of the generator. However, for rated generator load at minimum speed, the exciter voltage output and its required excitation are high, hence there are high losses; while at generator synchronous speed the exciter output and losses may both be low. Reasonably correct efficiency values may be hard to estimate for this important machine over the speed range. For the system indicated in Fig. 7, the statements are made in the paper that, "Here no frequency changer as such is employed. The field excitation of the exciter is at the differential frequency, so that it acts as a rotary amplifier." I cannot agree with this relatively simple conclusion since the exciter rotor or armature is rotating at an equivalent shaft frequency and under some conditions there may be present in the exciter output, the shaft frequency, the excitation or slip frequency, and their sum and difference. Some frequency-changing device is needed to obtain only the amplified slip frequence W_e-W_t for use of the generator rotor.

T. B. Owen: I believe the two questions asked by Mr. Hoard can be answered by one statement. The over-all efficiency curves of both Figs. 8(A) and (B) are based on three simple assumptions:

1. The generator efficiency K_g is 85% for any load put on it, at any input speed. Obviously, this can only be true for a limited area of loads, speeds, and power factors, but any other approach leads to utter confusion.

2. The differential loop efficiencies, K_1, K_fK_{sm}, K_{ss}, are not individually segregated. That is, for a constant differential efficiency curve of 0.9, say, the product $K_t \times K_f \times K_{sm}$ or $K_1 \times K_f \times K_{ss}$ is 0.9. The in-

dividual efficiencies can be anything, as long as the product is 0.9.

3. No account is taken of excitation. This is a loss which must be factored in, as well as windage, copper, iron, etc.

In order to define the limits of this paper, as requested by Mr. Hoard, I would like to note that all I have done is to present a method of analysis, and to show that all methods come down to the same thing, whether they function hydraulically or electrically. The way the curves of Figs. 8(A) and (B) should be used is to assume some speed range, and from Fig. 8(A), find the over-all efficiency of the constant-speed drive plus generator, assuming that the differential loop efficiency of the pump-motor combination is known. Then, for the same speed range, Fig. 8(B) will give the required differential loop efficiency for the systems of Figs. 6 and 7.

If the efficiency of the latter is appreciably less than that of the former, and shows the possibility of less complication, then it will be worth further investigation. I think that if Mr. Hoard uses my paper in this manner, he will not find very significant differences between the over-all efficiencies between the constant-speed and the constant-frequency systems.

I do not quite understand Mr. Hoard's statement regarding Fig. 7. If he will refer to Fig. 6, he will note that there is a physical, discrete, frequency changer between the exciter and the generator. In Fig. 7, a frequency changer, as such, certainly is not present. Further, any generator may be considered as a rotary amplifier; the output voltage, frequency, and power are functions of shaft speed, load, and field excitation. I believe my statement will stand as written.

I have serious doubts that the development of low-loss semiconductors will do much toward making the "differential" or "frequency make-up" systems practical. This, I will admit, is an unpopular opinion, since almost everybody is going in the other direction, but if the frequency converter of Fig. 6 and the rotary amplifier of Fig. 7 are practical for the applications shown, they should be even more practical when used directly as generators. For example, a 60-kva "frequency make-up" system will require at least a 20-kva, and possibly a 30-kva, frequency changer, and there are a number of generators of this capacity in use today. To my mind, this is a much more practical application than in the systems shown, and would be a major step toward simplifying systems now in use.

My thanks go to Mr. Hoard and to those who made oral comments on my paper. I trust that the paper will be helpful in evaluating the problems we have in this area. Also, I am very much indebted to Mr. Johnson of Hallamore and Mr. Chirgwin of Jack & Heintz for valuable discussions prior to writing this paper. Without their comments, I could not have prepared it.

Practical Considerations of an Ion Propulsion System

A. E. LENNERT
NONMEMBER AIEE

IT CAN logically be assumed that within 5 years it will be possible to project into orbit a 20,000-lb (pound) vehicle, five times as heavy as the largest satellite now in orbit. This payload capability is a necessary intermediate step in the long-range program leading to manned space flight.

Today's space vehicles cannot maneuver nor, until the recently announced SNAP III device, could they carry on uninterrupted communications for a great length of time. Control and communication then are the main obstacles blocking the way to space travel for other than exploratory and academic purposes.

The ion rocket is a reasonable system which has the inherent growth potential to power a space vehicle on extended journeys. In various forms, it has been the subject of several papers. This paper presents a parameter study of the ion propulsion system in terms of driving a 20,000-lb vehicle into space where the power source stems from the direct conversion of nuclear to electric energy.

Comparison of Space Propulsion Systems

An ion system may be justified mainly by the requirements of future space vehicles, requirements such as extreme reliability, development time cost, power source, specific power generation, specific impulse, payload capability, and many others too numerous to mention. Table I briefly outlines the characteristics of a number of propulsion systems that have been mentioned for use in space flight.

The ion system is among those that meet the major requirement for space propulsion: high I_{sp} and hence low w. Other systems with this quality present their own problems (see the Appendix which compares all the systems on a basis of thrust-to-weight ratio, $T/W = 5 \times 10^{-4}$, while carrying a payload of approximately 3,000 lb).

Ion Rocket Description

A conceptual design of a preliminary ion rocket system, which can be called a test bed, is shown schematically in Figs. 1 and 2. It functions in this manner. The cesium tank and associated pressurization and control equipment are in a compartment between the direct conversion nuclear-electric power supply and the d-c to d-c inversion-transforming housing. On either side of the cesium tank are thin sheets of highly reflecting material which prevent a large temperature rise within the tank and accelerating equipment compartments. With the aid of proper controls the cesium flows into the ion-forming region and is sprayed through carefully de-

Table I. Principal Propulsion Parameters

	Chemical	Plasma	Solar	Radio-isotope	Photon	Ion
Specific impulse, seconds (I_{sp})	400*	900 – 2×10^4	high	high	high	2×10^4
Thrust to weight, acceleration in g's (T/W)	2†	8×10^{-4}	low	low	extremely low	$10^{-4} \rightarrow 5 \times 10^{-4}$
Electric power thrust, kw/lb. (P_e/T)	25	30 → 300	zero‡	low	low	286 → 500

* Considered an upper limit of the chemical system.

† Considered minimum at take-off, increases with altitude in view of the decrease in gravitational field and atmospheric pressure.

‡ Depends upon the microscopic viewpoint that the sun's energy is being absorbed and then reradiated, for which $w \neq 0$, or macroscopically reflected for which $w = 0$.

Paper 59-833, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 20, 1959; made available for printing April 17, 1959.

A. E. LENNERT is with the Martin Company, Baltimore, Md.

The author wishes to express his appreciation to S. Taborsky for his suggestions and circuit design for the d-c to d-c conversion system and to L. D. Hassell for his assistance with the design of the cesium flow system.

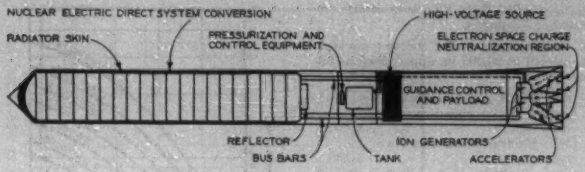


Fig. 1. Ion test bed layout

signed nozzles against a tungsten spray pan, see Fig. 2. When an atom of an alkali metal strikes a hot tungsten surface an electron is given up by the neutral atom and then rebounds in the form of a positive ion.

If the ionization potential, E_i, of the neutral impinging atom is lower than the work function, $e\phi$, of the material being struck, then there is a certain probability for the impinging atom to lose its valence electron. The ratio of the number of positive ions, n^+, evaporating from the surface of the tungsten spray pan to the number of neutral atoms, n, may be determined from the following relationship.[1]

$$n^+/n = e^{\frac{e\phi - E_i}{kT}}$$

where

$k = 8.5 \times 10^{-4}$ ev/K (electron-volts per degree Kelvin)
T = degrees K
E_i and $e\phi$ are measured in ev

As a typical example, consider the case of cesium being sprayed upon a hot tungsten surface maintained at a temperature of 1,200 K. $kT = 0.1$ ev; $e\phi - E_i = 4.52 - 3.87 = 0.65$. Thus, the ratio $n^+/n = e^{6.5}$, or approximately 400, results, which for all practical purposes indicates 100% positive ion formation. Experimental data verifying the analytical solution may be found in reference 2. The positive ions thus created are extracted from the ion-forming region by means of suitably designed electrostatic fields which control the drift of positive ions into an accelerator region. There, suitable potentials are applied to accelerate the positive ions to a predetermined velocity, yielding the thrust component for which the system is designed.

A separate, but similar, system is required for the removal, focussing, and acceleration of electrons; however, the accelerating voltage in this instance is extremely small compared with that required for the positive ions. This is due to the difference in masses of the electrons and the positive ions, clearly shown by equation 5. Instead of using a complete system to accelerate electrons, it is possible to bleed the excess electrons from the tungsten spray pan and, with sufficient potential, boil them off at the same velocity and direction as the positive ion stream thus neutralizing the positive ion space charge. This design feature is necessary in order to overcome the deleterious effect of space caused by the positive ion beam. In addition, it tends to keep the vehicle electrically neutral thereby preventing additional problems due to a highly charged vehicle.

Velocity components of the electrons should be comparable to the positive ion velocities so that the electrons will remain in the close proximity of the positive ions, hence increasing the probability for charge neutralization.

Mathematical Analysis

The equations for analysis of an ion propulsion system are interrelated in graphical form at the end of this section.

The space-charge-limited current density within a constant accelerating field of voltage, E_b, for the case of parallel plate geometry is described by reference 3.

$$j = \frac{i}{S^2} = \frac{\epsilon_0}{2.25}\sqrt{\frac{2eQ}{m_p M}}\,\frac{E_b}{d^2}\ \frac{\text{amp}}{\text{m}^2} \qquad (1)$$

where

i = ion beam current, amperes (amp)
ϵ_0 = permittivity in free space, meter-kilogram-second system
e = charge on the electron, coulombs
m_p = mass of the proton, kilograms
Q = degree of ionization (in this report singly only)
M = atomic weight of the ion (cesium 132.91)
E_b = accelerating voltage
d = plate spacing, meters
S^2 = area of the plates, meters^2

To obtain an expression for electric power P_e that is applied to the ion accelerator, as a function of vehicle thrust (T), it is noted that

$$P_e = E_b i \qquad (2)$$

and from fundamental considerations

$$i = 0.45 \dot{w}\,\frac{eQ}{m_p M}\ \text{amp} \qquad (3)$$

where $\dot{w}$ is the propellant flow rate in lb/-sec (second). The thrust is obtained from Newton's second law:

$$T = d/dt\left(\frac{w}{g}v_e\right) = \frac{\dot{w}v_e}{g}\ \text{lb} \qquad (4)$$

In this equation the exhaust velocity of the ions is assumed to be constant, which is quite valid so long as the accelerator plate potential is constant.

g = acceleration due to gravity, meters/sec^2
v_e = exhaust velocity of the accelerated ions

Exhaust velocity is described by the following equation:

$$v_e = \sqrt{\frac{2eQ}{m_p M}E_b}\ \text{meters/sec} \qquad (5)$$

Combining equations 2–5 gives the power required per unit thrust

$$\frac{P_e}{T} = 30.8\sqrt{\frac{QE_b}{M}}\ \text{kw/lb} \qquad (6)$$

(Equation 6 leads to a conservative estimate of the thrust for given electric power because, in equations 4 and 6, acceleration due to gravity (g) is assumed constant. Actually, in a 2-body problem, it decreases with altitude as $g = g_s(r_s/r)^2$, g_s being the value at the earth's surface, 9.8 meters/sec,^{2} and r_s being the radius of the earth and $r = r + h$ where h is the height above the earth's surface.)

Another useful expression, relating the thrust and the propellant flow rate, from equations 4 and 5

$$T = \frac{1}{g}\sqrt{\frac{2eQ}{m_p M}E_b}\,\dot{w}\ \text{lb} \qquad (7)$$

The design of the ion accelerator portion itself requires a knowledge of the voltage gradient, area of the plates, and the thrust demanded by the system.

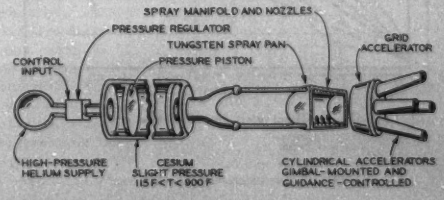


Fig. 2. Cesium system layout

These are obtained from equations 1, 3, 4, and 5 in the form. From equation 8, given a value of the voltage gradient E_b/d in volts/meter, the area of the accelerator plates required for each pound of thrust may be obtained.

$$\frac{S^2}{T} = \frac{6.03 \times 10^{\sigma}}{(E_b/d)^2} \text{ ft}^2 \text{ (square feet)/lb} \qquad (8)$$

Equations 1 through 7 consider the major parameters required for the study of the ion system itself, with no regard for the vehicle trajectory in space. However, such quantities as thrust, T, and over-all weight, W, do influence the vehicle trajectory and are particularly important for any contemplated mission. Indeed, the ratio of thrust to weight is extremely important for this is the g acceleration the vehicle is constantly experiencing, the key factor in determining the time required to complete any mission.

Graphical Interrelationship of Parameters

The ion system parameters just described in equation form are presented graphically in Figs. 3–8. From these curves a conceptual design of the ion vehicle can be projected, based upon the following assumptions:

1. The vehicle is initially in the earth's orbit, projected there by a booster mechanism comparable to the advanced Titan system.

2. Minimum perigee is approximately 300 miles. (This condition is imposed because drag forces are extremely low at this altitude with negligible effect on the applied thrust. Reference 5 reports that a vehicle, with area dimensions somewhat greater than the current design, experienced a drag force of approximately 10^{-6} lb at 300-mile altitude.)

3. Controlled flight time is arbitrarily selected as 500 days.

4. Total vehicle weight (booster payload) is selected as 20,000 lb, apportioned as follows:

(*a*). Direct nuclear-electric reactor and associated control equipment	8,000
(*b*). Electric transformation equipment	2,000
(*c*). Tankage and allied equipment	1,000
(*d*). Ion fuel (based on cesium which imposes the greatest weight penalty)	3,000
(*e*). Accelerator and allied equipment	1,000
(*f*). Structure	1,000
(*g*). Payload (guidance and control equipment, cameras, transmitters, receivers, scientific instruments, etc.)	3,000
Subtotal	19,000
Contingencies	1,000
Total	20,000 lb

5. Power at the bus bar is 400 kw.

Fig. 3 shows the variation of thrust with electric power, using the accelerator plate voltage and the ion propellant as parameters. It is readily seen that for a given power setting thrust varies inversely with plate voltage and directly with the atomic weight of the propellant. It appears, at the outset, that if power is the overriding concern, the accelerator plate voltage should be as low as possible and materials having the greatest atomic weight should be selected.

Fig. 4 is a plot of thrust versus propellant flow rate, with the parameters again being the accelerator plate voltage and the atomic weight of the propellant. If propellant weight is the dictating parameter, then it appears that the propellant should be of the lowest atomic weight.

Fig. 5 indicates thrust versus ion current, with the accelerator plate voltage and the propellant as additional parameters. For a given thrust and plate voltage, it is seen that the ion current varies inversely as the atomic weight of the propellant. For a given plate voltage, the exhaust velocity of the ion beam increases inversely as the square root of the atomic weight of the material as described by equation 5.

Fig. 6 (obtained from equation 3) describes the variation of propellant flow rate with ion current for the two materials being considered, cesium and potassium.

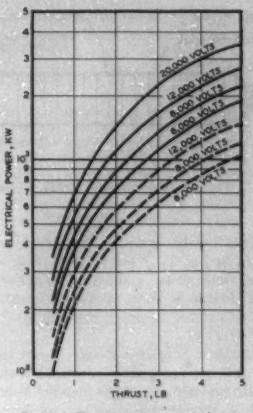


Fig. 3. Electric power versus thrust

– – – –Cesium
———Potassium
Voltage parameter is the accelerator plate voltage

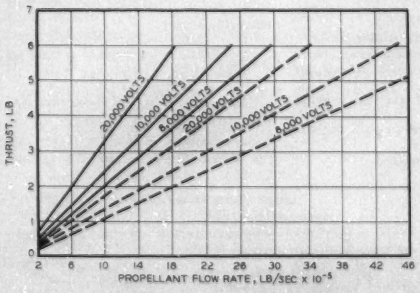


Fig. 4. Thrust versus propellant flow rate

– – – –Cesium
———Potassium
Voltages are the accelerator plate voltage

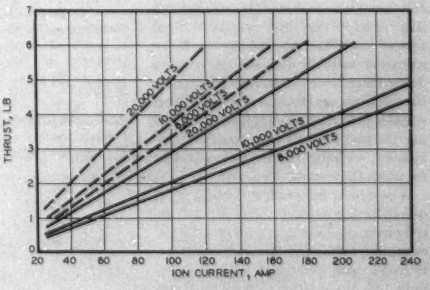


Fig. 5. Thrust versus ion current

– – – –Cesium
———Potassium
Voltages are the accelerator plate voltage

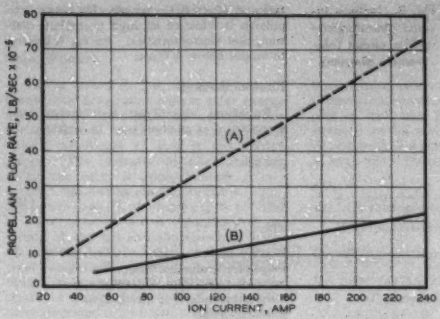


Fig. 6. Propellant flow rate versus ion current

----Cesium
——Potassium

The curves show that the ion current required for a given flow rate decreases as the atomic weight of the propellant increases. Again, if propellant weight is a criterion, then the highest weight should be sought, consistent with the dictating requirement of producing positive ions efficiently.

Fig. 7 summarizes a number of accelerator characteristics. Plate spacing between the accelerator plates is plotted against accelerator voltage. Parameters are the field strength (volts/meter) and the surface area required per unit thrust (ft/lb). For a field strength of 10^8 volts/meter, accelerator spacing is in the order of 10 cm (centimeters) for 12,000 volts on the accelerator plates. Although this design feature is quite practical, the accelerator surface area required, because current density is limited by the effects of space charge, is approximately 600 ft^2/lb thrust. This is rather an impractical value for the present system design. Increasing the field strength by an order of magnitude to 10^8 volts/meter indicates that the plate spacing would be in the order of 1 cm while the required plate area is reduced by approximately 6 ft^2/lb of thrust. As these values are reasonable and practical to attain, they merit further consideration in the system design.

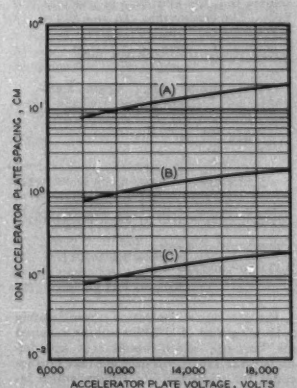


Fig. 7. Plate spacing versus plate voltage

A—$E_b/d = 10^8$ volts/meter; $S^2/T = 603$ ft^2/lb
B—$E_b/d = 10^8$ volts/meter; $S^2/T = 6.03$ ft^2/lb
C—$E_b/d = 10^7$ volts/meter; $S^2/T = 0.06$ ft^2/lb

Upon increasing the field strength still further, it is seen that the accelerator plate spacing is in the order of 0.1 cm and the accelerator plate area in the order of 0.06 ft^2/lb of thrust. Although the latter value is quite attractive and appears practical, the small plate spacing poses a number of problems. In this region, the field is strong enough to create disturbances within the ion beam. Electrical breakdown may result, thereby destroying the entire accelerator system. Therefore, the field strength should be considered initially in the order of 10^8 volts/meter with a minimum of 12,000 volts on the accelerator plates. The other values should be adjusted accordingly.

The reason for selecting 12,000 volts as the minimum plate voltage is clearly shown in Fig. 8, a plot of ion beam and accelerator plate currents as a function of accelerator plate voltage. With relatively low voltages applied to the accelerator plates, the plates tend to draw at least an equal, if not greater, amount of current than that of the ion beam. This phenomenon occurs because, when the accelerator voltage is relatively low, a considerable portion of the ion beam current is attracted to the accelerator plates. Fig. 8 indicates that the accelerator current reaches its maximum at 5-kv plate potential and then decreases markedly as voltage is increased. The accelerator beam current becomes asymptotic in the neighborhood of 12 to 14 kv. It is believed this large value for the accelerating potential could be decreased by designing guard rings properly and by cooling the acceleration plates.

Electrical System

The electrical system functions in the following manner. The output from the nuclear reactor is in the form of high-current–low-voltage direct current. By means of properly designed busbars the current flows to the inversion-transforming region. The output from the nuclear power plant is in the order of 500 kw. With approximately 80% over-all efficiency, the power to the accelerator is approximately 400 kw.

The inversion process may be effected with the use of "gated" silicon rectifier cells in a manner similar to the ignitron conversion, with the gate serving as a commutation device used in conjunction with a sequence timer. Conventional silicon cells currently in operation may carry peak currents of 1,000 amp with the gated version available to carry approximately 25 amp. With several years of further developmental effort, larger ratings for this application may be anticipated in a low-duty-cycle arrangement as shown in Fig. 9. Frequency of operation would be limited by the permissible duty cycle.

The voltage drop across the silicon cells is approximately 1 volt with a resultant efficiency of approximately 98% for this application. The rectifier requirement is a half-wave 6-phase circuit with a number of silicon rectifiers in each phase depending upon the ratings of the rectifiers. The over-all efficiency from the reactor output to the bus bars is about 80%.

Conclusions

As in many other types of propulsion, there are a number of unsolved problems

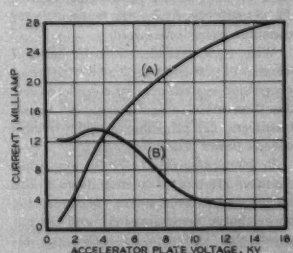


Fig. 8. Current versus accelerator plate voltage

A—Ion beam current
B—Accelerator electrode current

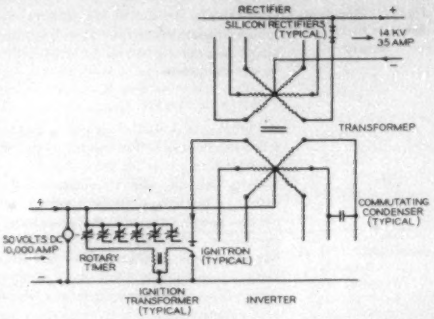


Fig. 9. Ignitron inverter - transformer - rectifier circuit (elementary diagram)

inherent in ion propulsion. The ultimate objective is a space vehicle system of optimum efficiency but, at first, efficiency itself is not the paramount criterion from which all other design considerations must originate. It is of more immediate importance to design, construct, and ground-test an operational vehicle and then optimize the system for the greatest over-all efficiency. This is particularly true of the ion system where a strong interrelationship exists among parameters. Individual parameters might be superficially optimized only to find that the process upsets some other paramerer.

As an example, it is sometimes stated that the greater the plate voltage, the greater the exhaust velocity, and hence the more efficiency the system. That this is not necessarily true can be seen from equations 3, 4, and 5, which indicate that an increase in accelerator plate voltage E_b increases the exhaust velocity v_e of the positive ions. For constant power, an increase in plate voltage E_b corresponds to a decrease in the current i. However, the thrust T depends more strongly upon the current than upon the voltage. In fact, an increase in accelerator voltage E_b actually decreases the thrust under constant power operation.

With these considerations in mind, a number of problems associated with anion propulsion system may be considered under three main categories: (1) ion thrust chambers, (2) propellants, and (3) high specific power generation.

Problems associated with the ion thrust chamber are related to such items as the ion source, electron source (necessary for space charge neutralization), and the ion accelerator system. In each case, although design feasibility may be established by means of the equations, further theoretical and experimental work, in the tens of ampere range, is essential. Increasing the voltage gradients raises the space-charge-limited currents which, for constant current, implies reducing the accelerator plate areas. Detailed experimental effort is obviously in order.

With regard to propellants, the over-all propellant weight for an alkali ion system appears to vary between limits that are not too great, whether cesium or sodium is used. That is, alkali metals have atomic weights between 23 and 133, so that at constant current the propellant flow rate, and hence total propellant weight W_p, may vary by a factor of 6 (equation 3). It may be possible to utilize heavy molecules for the propellant, such as SbI_3 which has a molecular weight of approximately 757, the provision being that it would be possible to remove an electron. This would bring about a new positive charge on the molecule and enhance the system's characteristics. This area merits more detailed effort.

The third category, high specific power generation, also requires considerable experimental effort in building a direct nuclear-electric conversion system as outlined in Fig. 1. At present the fuel loading appears enormous (approximately 20 lb per linear ft of reactor) and the power density is on the order to 35 kw per cubic ft. In addition, further work is needed on semiconductor rectifiers to handle extremely high currents and voltages.

There is no doubt that the problems inherent in the ion system design challenge the engineering mind. However, it is strongly believed that if the requirement for such a system were generated these major problems would be resolved.

Appendix. Comparison of Space Propulsion Systems[6]

Chemical Rocket

Even though the chemical rocket exhibits high thrust and a high thrust to weight ratio, it offers little promise for space propulsion in view of its high propellant consumption characteristics. Its I_{sp} is limited to values below 400 sec.

Photon Rocket

Equations describing the energy and momentum of photons may be written:

$$E = h\nu = \frac{hc}{\lambda} \tag{9}$$

$$p = h/\lambda \tag{10}$$

where

h = Planck's constant
ν = photon radiation frequency
c = velocity of light 3×10^{10} cm/sec
λ = photon radiation wavelength
p = momentum

Since the thrust

$$T = d/dt(P) \tag{11}$$

by combining equations 9–11 the following is obtained:

$$T = d/dt(E/c) = \frac{1}{c}\frac{dE}{dt} \tag{12}$$

A photon rocket, depending upon a nuclear reactor for its source of photons, will be assumed to weigh approximately 20,000 lb with a 3,000-lb payload. For T/W of 5×10^{-4}, the thrust would be approximately 1 lb. For 1 lb of thrust equals 4.45×10^5 dynes, the fuel burn-up required per unit time is obtained by the use of equation 12:

$$\Delta E = \int \frac{dE}{dt}\,dt = \int cT\,dt = cT\Delta T$$
$$= 3\times10^{10}\times4.45\times10^5$$
$$= 8.34\times10^{21}\ \text{mev/sec} \tag{13}$$

Consider the use of fissionable material (U-235) for the creation of photons and assume that the efficiency of production is 10%. The energy that can be supplied

$$E_s = (0.10)200\ \frac{\text{mev}}{\text{fission}} \times 6.02\times10^{23}\ \frac{\text{atoms}}{\text{gm(gram)-atom}} \times \frac{1}{235}\ \frac{\text{gm-atoms}}{\text{gm}} \times 454\ \frac{\text{gm}}{\text{lb}}$$
$$= 2.33\times10^{25}\ \frac{\text{mev}}{\text{lb U-235}} \tag{14}$$

Dividing equation 13 by 14, it is seen that

$$\frac{8.35\times10^{21}\ \text{mev/sec}}{2.33\times10^{25}\ \text{mev/lb of U}} = 3.57\times10^{-3}$$
$$\text{lb/sec} = 1.63\ \text{gm/sec} \tag{15}$$

For one day of operation approximately 1.41 kilograms of U-235 would be required for burn-up, which would cost approximately $21,125. This does not take into account the nonfission absorptions which increases the fuel investment by approximately 10%, nor is the normal fuel loading required for criticality considered. The fuel required for burn-up due to all processes for long-term operation is only about 20% of the total fuel investment. Fuel cost

alone makes the photon rocket unattractive now.

Solar Rocket

A solar rocket derives its component of thrust from a sail that reflects solar energy (photons). The sail must have considerable surface area in order to reflect a sufficient amount of energy to give it the desried thrust. The intensity of solar radiation about the earth is approximately 1 kw/meter^2[6] and on reflection from the sail (say, aluminum-coated), the intensity is doubled to 2 kw/meter^2. If it is assumed that the weight of the sail and the structure is approximately half the vehicle weight, then a total weight $W=6{,}000$ lb and the sail weight is approximately 3,000 lb. For the thrust to weight ratio T/W of 5×10^{-4} the thrust is found to be 0.3 lb. Using equation 12, $dE/dt=4\times10^8$ kw, the sail area required for this thrust is

$$\frac{4\times10^8 \text{ kw}}{2 \text{ kw/meter}^2}=2\times10^8 \text{ meter}^2$$

This is an area, if taken as a square, with a side of approximately 480 yards. Also, in view of its extremely small skin thickness, the solar sail is subject to rupture by penetration of meteorite particles and does not appear practical for long range space flight, nor does it have a promising growth potential for increase in power density.

Radioisotope Rocket

Space propulsion may be accomplished by means of a polonium sail[7] which utilizes the natural radioactive decay particles of the isotope Po-210. Po-210 decays by alpha-particle emission, and a sail coated with a thin layer of the material could provide trouble-free operation. Other alpha-particle emitters are also possible, although Po-210 seems to strike a good compromise between alpha-particle decay energy and relatively long half-life.

Three parameters of interest for this Po-210 sail system are (1) dynes thrust/gm of Po-210, (2) areal density (gm/cm^2), and (3) cost. To obtain parameter 1, note that the thrust/gm of Po-210 is

$$T=\frac{d}{dt}(mv)\doteq v\frac{dm}{dt}=v\frac{dN}{dt}(m_a) \qquad (16)$$

where

m = total mass ejected from rocket
m_a = mass of alpha-particle
v = velocity of alpha-particle
dN/dt = number of Po-210 atoms disintegrating/sec/gm of Po-210 with emission of an alpha-particle, i.e, number of alpha-particles/sec/gm of Po-210

Po-210 emits an alpha-particle of average energy $E=5.3$ mev, so that

$$v=\sqrt{\frac{2E}{m_a}}=\sqrt{\frac{2(5.3)1.602\times10^{-6}}{(4)\,1.66\times10^{-24}}}=1.6\times10^8 \text{ cm/sec} \qquad (17)$$

Recent information from an Atomic Energy Commission representative at Mound Laboratory indicates Po-210 with a source strength of 1.4 micrograms/millicurie may be obtained at a cost of $0.10 per millicurie in large lots. Therefore, for 1 gm of Po-210

$$\frac{dN}{dt}=\frac{1}{1.4 \text{ microgram/millicurie}}=(715)\ 3.7\times10^{10} \text{ alpha-particles/sec/gm Po-210} \qquad (18)$$

Combining equations 16–18

$$T=1.6\times10^8(715)\,3.7\times10^{10}(4)\,1.66\times10^{-24}=0.28 \text{ dyne/gm Po-210}=6.3\times10^{-7} \text{ lb thrust/gm Po-210} \qquad (19)$$

The areal density (gm/cm^2) of the Po-210 coating on the sail can be conservatively estimated by choosing a thickness of Po-210 equal to the range of 5.3 mev alpha-particles in polonium itself. Since this range turns out to be about 10^{-4} cm, and the volume density of polonium $\rho=9.4$ gm/cm^3, the areal density becomes 9.4×10^{-4} gm/cm^2.

To determine the third item (cost), it is necessary to know the sail area required. Here again a conservative guess of sail weight equal to half the total vehicle weight implies, for a 3,000-lb payload, a sail of about 3,000 lb, and total vehicle weight of about 6,000 lb. But, with $T/W\sim5\times10^{-4}$ and $W=6{,}000$ lb, $T=0.30$ lb$=1.3\times10^6$ dynes. This, coupled with equation 19, indicates that the amount of polonium needed is

$$(1.3\times10^6 \text{ dynes})/(2.8\times10^{-1} \text{ dynes/gm Po-210})=4.65\times10^6 \text{ gm Po-210} \qquad (20)$$

From equation 18, 715×10^3 millicuries/gm Po-210 is the available source strength, which results in a total strength of

$$(4.65\times10^6 \text{ gm Po-210})(7.15\times10^5 \text{ millicuries/gm Po-210})=3.32\times10^{11} \text{ millicuries} \qquad (21)$$

At a cost of $0.10 per millicuries this means that the necessary sail would cost about 33 billion dollars, which is, of course, the most significant deterrent to the whole concept.

The actual sail size, which depends on the areal density and the mass of Po-210, is then

$$(4.65\times10^6 \text{ gm})/(9.4\times10^{-4} \text{ gm/cm}^2)\rightarrow 2{,}420 \text{ yards}^2 \qquad (22)$$

An equivalent square would thus have a length of 49.3 yard on each side.

Finally, it must be borne in mind that although the 33.2 billion dollar cost is prohibitive, it is based on a 3,000-lb payload with $T/W\sim5\times10^{-4}$. For smaller payloads, it may well be that the radioisotope sail is more attractive. However, one drawback is the difficulty of both obtaining the needed Po-210 and coating the sail.

Plasma Rocket

W. Bostick states that his plasma gun fires about 10^{-7} gms of plasma per pulse, and that velocities of the order of 10^8 cm/sec can be obtained for these plasmas. Since Bostick visualizes the need for a primary electric power source for his plasma gun rocket, as, for example a nuclear reactor system, it is reasonable to assume, as for the ion system, a total vehicle weight $W=20{,}000$ lb. Maintaining the T/W ratio at 5×10^{-4} implies a thrust of $T=[1.0 \text{ lb}]=4.45\times10^5$ dynes. In general though, $T=(4.45\times10^5)/10^7=4.45\times10^{-2}$ gm/sec. Each gun is capable of 10^{-7} gm plasma/pulse, yielding 0.45×10^6 pulses/sec, or about 2 microsec/pulse.

If this can be accomplished by one gun, or perhaps several guns operating together, then the plasma system is a promising one.

An inherent advantage of plasma propulsion over ion propulsion is the essentially neutral character of the plasma. That is, while space charge neutralization is necessary for the ion system, no such provision must be made for the plasma system. However one disadvantage lies in the electrode erosion[8] during continuous operation which tends to limit the use of plasma systems.

In conclusion, Bostick indicates that current work on plasma acceleration by so-called rails, etc., may soon allow plasma velocities to become as high as 10^8 cm/sec. This is a significant advance, at least as regards propulsion possibilities.

References

1. Electronic and Ionic Impact Phenomena, H. S. W. Massey, E. H. C. Burhop. Clarendon Press, Oxford, England; also Oxford University Press, New York, N. Y., 1952.

2. Ionization on Platinum and Tungsten Surfaces, I—The Alkali Metals, Sheldon Datz, E. H. Taylor. *Journal of Chemical Physics* New York, N. Y., vol. 25, no. 3, Sept. 1956.

3. Electronics, J. Millman, S. Seely. McGraw-Hill Book Company, Inc., New York, N. Y., 1941. p. 211.

4. Ion Source for the Production of Multiply Charged Heavy Ions, C. F. Anderson, K. W. Ehlers. *Review of Scientific Instruments*, American Institute of Physics, New York, N. Y., vol. 27, no. 10, Oct. 1956, pp. 809–15.

5. American Physical Society, New York, N. Y. May 1958.

6. Ion Rocket Test Bed, P. E. Grant, A. E. Lennert. *Engineering Report 9963*, Martin Company, Baltimore, Md., June 1958.

7. R. W. Bussard. *Jet Propulsion*, New York, N. Y., vol. 28, 1958, p. 223.

Discussion

R. W. Stineman (Boeing Airplane Company, Seattle, Wash.): The author's equation 5 results in an overly optimistic value for the exhaust velocity to be expected from an ion accelerator, because the equation implicitly assumes that all of the electric energy being supplied to the accelerator is converted to ordered kinetic energy of the exhaust stream. This may be a reasonable assumption for a few widely separated ions. However, in the relatively dense ion streams envisaged for propulsion apparatus, interaction between ions will partially disorder the motion of the ions, thus converting some of the energy into heat. Another factor is that an ion cannot fall through the full potential difference between electrodes, E_b, unless it actually strikes the attracting electrode. Ions which are discharged into space can be accelerated by only a fraction of the total voltage. Good

design can make this fraction large, but not unity. In general then, a factor should be included under the radical in equation 5 to represent the efficiency of converting electric energy to ordered kinetic energy. The determination and optimization of this efficiency is one of the problems to be solved.

Exception is taken to the author's contention that equation 6 is conservative. There appears to be a misinterpretation of the meaning of g in equation 4. Equation 4 is simply a statement of Newton's second law of motion. However, with force expressed in pounds, mass in pounds, and velocity in meters per second, a unit conversion factor is required. The fact that this constant unit conversion factor happens to be equal to the sea-level acceleration of gravity is coincidental. Note that the author's reasoning would lead to the untenable conclusion that in interstellar space, where local gravity is virtually zero, an infinite amount of thrust could be attained from a minute amount of power. In addition, the preceding discussion of equation 5 is equally applicable to the equations which are derived from it.

Figs. 3 and 4 present useful design information, but do not support the author's comments regarding selection of propellant. Any desired ratio of power to thrust or thrust to flow rate can be attained for any propellant material by selecting the accelerating voltage accordingly. The reason for selecting a heavy ion is best illustrated by Fig. 6, which shows that less current is required for a given flow rate with heavy ions, thus minimizing space charge problems.

Very general analyses of propulsion mechanics made by D. B. Langmuir and J. H. Irving of Ramo-Wooldridge have shown that minimization of initial gross mass depends only upon proper choice of exhaust velocity and ratio of power plant mass to propellant mass. Since the optimum exhaust velocity can be attained with any propellant material by proper selection of accelerating voltage, it is not clear why the author's weight analysis table implies a weight penalty for the use of cesium, especially when cesium tends to have fewer problems with space charge. It might also be pointed out that the afore-mentioned analyses show that when a 20,000-lb vehicle has a payload of 3,000 lb, then the optimum allocation of the remaining weight is about 12,500 lb of propellant and 4,500 lb for the power plant. The author's choice of 3,000 lb of propellant thus appears to be somewhat removed from optimum.

The source and derivation of Fig. 8 are not clear. It seems likely that the specific values shown are for some specific configuration, rather than having general applicability. Consequently, it does not appear that 12,000 volts is necessarily the minimum practical voltage, irrespective of electrode shape or spacing. Clarification of this point by the author would be appreciated. Also, it would be of value to have a further discussion of what is meant by electrical breakdown of the ion stream in a strong field, since it is not clear how already ionized gas would break down.

A. E. Lennert: Mr. Stineman raises some noteworthy points. Certainly the electrical-mechanical efficiency conversions would have to be considered as he suggests. They were not specifically mentioned in the paper because they are strongly dependent on the specific mission, which was not treated. Even so, the efficiency of conversion was taken into account in the analysis. It was taken to be approximately 80% with an additional 5% loss due to other items such as line losses, etc.

My usage of g should have been more correctly identified as the gravitational constant which, as Mr. Stineman states, happens to be the sea-level acceleration of gravity.

I must take exception to the remarks on the allocation of weights. Granting that propulsion mechanics may dictate optimum performance, one must nevertheless consider the practical aspects of the system design. Since the aggregate weight of the nuclear power plant plus controls, instrumentation, associated transformation equipment, skin, structure, and a host of other items amounts to 14,000 lb, 50% of the remaining weight (6,000 lbs) was reserved for payload. If, instead of 20,000 lbs, the boost capability within the time limit was increased to 100,000 lbs, then one could reasonably increase the propellant fraction to the amount dictated by theoretical propulsion mechanics. Here again optimum performance must be related to a specific mission. Since no mission was specified, we may consider that the vehicle is designed for rocket test purposes.

Anderson and Ehlers[1] have conducted experiments whereby one may establish preliminary estimates of the minimum voltage required for accelerating heavy ions. In the paper the ion beam passed through an electrode with a slotted aperture which was designed to satisfy Pierce's condition.[2] Phenomenologically the events were briefly summarized in my paper, however, further details may be obtained from references 1 and 2.

Finally, singly ionized cesium atoms may be ionized doubly, triply, and so on, until eventually all of the orbital electrons are removed from the atom. However, the term "breakdown of the field" implies something else, namely: as the field strength is continously increased, electrons are drawn from the loosely bound conduction electrons of the accelerator; also, other mechanisms, such as secondary electron emission, thermionic emission, etc., may cause electrons to enter the ion stream, thereby creating a conduction path resulting in flashover or breakdown, of the accelerating field. Further details may be found in reference 3.

REFERENCES

1. See reference 4 of the paper.

2. RECTILINEAR ELECTRON FLOW IN BEAMS, J. R. Pierce. *Journal of Applied Physics*, New York, N. Y., vol. 11, Aug. 1940, pp. 548–54.

3. HANDBUCH DER PHYSIK, VOL. 22, GAS DISCHARGES (book). Springer-Verlag, Berlin, Germany, 1956.

Optimum Reflector-Absorber Geometry for a Solar Generator

R. W. STINEMAN
ASSOCIATE MEMBER AIEE

AS AN initial step toward the eventual realization of a completely static system for converting solar radiation to electric power, the analysis and testing of a laboratory thermoelectric generator has been undertaken. A description of the over-all project is given in reference 1. An essential part of the over-all system is the combination of reflector and absorber used to collect and concentrate the solar radiation. This paper deals with the geometrical optimization of the reflector and absorber.

The elements of a power-conversion system must be treated somewhat differently from the analogous parts of a solar furnace. These differences arise from the following considerations:

1. The reflector and absorber in a power-conversion system must collect and transmit solar energy to the generating elements with an acceptable efficiency. Any reflected energy which does not impinge on the absorber, or which is reflected or reradiated from the absorber, must be considered as loss in the determination of efficiency. In contrast, a solar furnace operates (in the foregoing sense) at zero efficiency, with maximum absorber temperature being the prime objective.

2. Since the study of a power-conversion system is directed toward an ultimate installation on a space vehicle, it is reasonable to assume that irregularities in the reflector surface and misalignment of the reflector (with respect to the sun) may cause the reflected rays to deviate from the ideal. On the other hand, most solar

Paper 59-859, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing April 28, 1959.

R. W. STINEMAN is with the Boeing Airplane Company, Seattle, Wash.

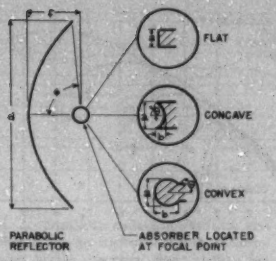


Fig. 1. Basic reflector and absorber geometry

furnace analyses have assumed a mechanically perfect reflector.[3-4]

Definition of the Problem

Briefly stated, the problem under consideration is to find the geometrical properties of the reflector and absorber which result in the most efficient collection of solar radiation for any given set of conditions. Collection efficiency is defined as follows:

$$\text{Efficiency} = \frac{\text{Net power absorbed}}{\text{Solar power intercepted}}$$

"Net absorption" refers to the power available for useful purposes after accounting for all losses, such as reradiation from the absorber surface. "Solar power intercepted" refers to all solar radiation directed toward the area enclosed by the perimeter of the reflector, regardless of whether such radiation actually reaches the reflector or is intercepted by an obstruction. The variables which enter into the determination of efficiency are now discussed.

Shape. At present, analysis has been confined to the cylindrical shapes shown in cross section in Fig. 1. The reflector is a parabolic cylinder. Three different absorber shapes are considered. The flat absorber has the advantage of minimum reradiating area for a given width. The circular cylindrical absorbers (especially the convex type) are more suited than the flat absorber for receiving radiation from a wide-angle reflector; i.e., when the angle ϕ shown in Fig. 1 is large.

Reflector diameter, d. With increasing reflector diameter, more radiation is collected but the reflected rays approach the absorber from a wider angle.

Tolerance on reflected rays, δ. As the result of manufacturing irregularities in the slope of a reflector and misalignment of the reflector-absorber assembly with respect to the sun, any given reflected ray may deviate from its ideal direction by an angular tolerance equal to $\pm\delta$. It may be noted that a manufacturing irregularity which produces a given angular perturbation of the reflector surface will cause the reflected ray to deviate by twice as great an angle. On the other hand, misalignment of the assembly by a given angle will cause a reflected ray to deviate by an equal angle (relative to the absorber), since any misalignment results in a partially compensating relative rotation of the reflector and absorber. Linear (as opposed to angular) manufacturing irregularities are disregarded, since their effects would be negligible compared with angular irregularities.

Maximum allowable angle of incidence at absorber, γ. Certain selective absorbing surfaces, to be discussed later in more detail, require for efficient performance that radiation be incident somewhere near perpendicular to the surface. The angle γ is the maximum allowable deviation from perpendicular.

Absorber width, a. The minimum absorber width which will intercept all of the reflected radiation with an acceptable angle of incidence depends on all of the previously discussed variables; i.e., absorber shape, d, δ, and γ. If all variables but d remain fixed, then as d increases, a must also increase. The increase in d increases the amount of solar radiation being collected, but the corresponding increase in a results in increasing losses. If a point is reached where the losses increase proportionately faster than the radiation collected, then an optimum condition will be reached. Subsequent analysis will show that such an optimum does exist at which efficiency is maximized.

Absorber depth, b. The shape of a circular cylindrical absorber may be quantitatively specified by means of the distance b, shown in Fig. 1. Large values of b are most suitable for wide-angle reflectors. Subsequent analyses employ the ratio b/a, rather than b alone.

Reflector focal length, f. By expressing the other lengths shown in Fig. 1 in terms of a ratio of f, the focal length is eliminated as an explicit variable. This is essentially the same as fixing f at a value of 1 unit. If the other lengths remain at fixed proportions of f and f varies, then both the net power absorbed and solar power intercepted will vary proportionately, and efficiency will be unchanged. Thus, the efficiency obtained is valid for any value of f.

Corrected solar radiation intensity, I. The intensity of solar radiation is the power-per-unit area normal to the sun's rays. The intensity after reflection is somewhat reduced since no reflector is perfect. However, for purposes of analysis, it is acceptable to assume a perfect reflector and to reduce the value of the intensity to account for the actual reflectivity. Thus, corrected intensity, I, represents the actual solar intensity multiplied by the specular reflectivity of the reflector. Since the corrected intensity is used in the determination of solar power intercepted, the calculated efficiencies do not include any effect due to losses at the reflector.

Absorber temperature, T. The analysis is performed in terms of a specified absorber temperature, rather than finding the temperature for a specified configuration, both for ease of analysis and because the requirements of the generating elements will usually require a specified input temperature. Since reradiation losses from the absorber depend on its temperature, it may turn out for a given set of conditions that the net power absorption is negative. This would simply mean that the specified absorber temperature is not realizable with the other given conditions, even at zero efficiency. It is apparent that the absorber temperature is not entirely arbitrary, but must be kept within reasonable bounds in order to obtain realistic results.

Selectively absorbing surfaces. Several publications have described developments of surfaces which are good absorbers of wavelengths in the high-energy part of the solar spectrum, but are poor emitters in the high-energy part of the infrared radiation spectrum characteristic of bodies at moderately elevated temperatures.[6-8] For such surfaces, the over-all effective absorptivity for solar radiation is denoted α, while the over-all effective emissivity at the specified absorber temperature is denoted ϵ.

Analysis and Results

Details of the method of analysis are given in the Appendix. The required calculations are not complex. However, the number of variables is such that several dozen cases must be considered to even roughly determine the effects of all variations. Consequently, it was economical to execute the calculations on an IBM (International Business Machines Corporation) 704 digital computer. Details of the computer program are not given, but should be obvious to any reader with computer experience. The

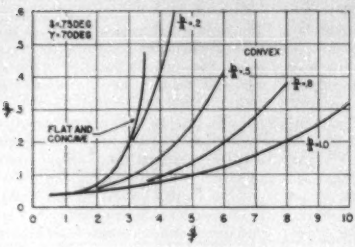


Fig. 2 (left). Typical variation of absorber width versus reflector diameter

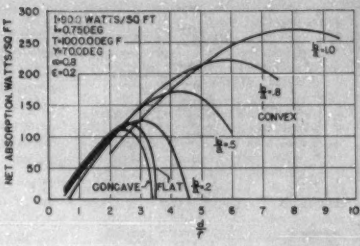


Fig. 3 (right). Typical variation of net power absorption versus reflector diameter

essential nature of the analysis is outlined as:

1. The radiation reflected from an element of the cylindrical reflector falls within a wedge having an angular width which exceeds the angular diameter of the sun by 2δ, where δ is the angular tolerance of reflected rays. Assuming the radiation to be uniformly distributed within the described wedge, it can readily be shown that collection efficiency is maximum if the absorber is just large enough to intercept all of the reflected radiation. The ratio a to f (see Fig. 1) can then be determined as a function of d/f if the absorber shape and the angles δ and γ (see preceding section) are specified. Curves of a/f versus d/f are shown in Fig. 2 for typical values of δ and γ.

2. The solar power intercepted for unit length (1 foot) perpendicular to the plane of Fig. 1, is the product of I and d, which, when expressed per unit (pu) focal length, becomes Id/f. The solar power reaching the reflector and subsequently impinging upon the absorber is somewhat less because some of the radiation is blocked by the power-conversion equipment installed adjacent to the absorber. The width so blocked is assumed to be equal to the absorber width.

3. After obtaining the radiation impinging upon the absorber, multiplying by the absorptivity, α, and subtracting the reradiation loss yields the net power absorption. Then dividing by the solar power intercepted gives the collection efficiency. Curves of net power absorbed versus d/f for a typical condition are shown in Fig. 3. Fig. 4 shows curves of collection efficiency versus d/f for the same condition as Fig. 3. The efficiency is independent of the focal length, f, and the distance perpendicular to the plane of Fig. 1.

Selective Surfaces

In all cases investigated, the use of a selectively absorbing surface resulted in higher efficiency than with a black surface. For example, with all conditions the same as in Fig. 4 except for changing the absorber surface to black (i.e., $\alpha=\epsilon=1.0$), the maximum efficiency is only 0.07 pu. (As is shown in the Appendix, a black absorber is superior to any other surface with constant emissivity.) The superiority of selective surfaces undoubtedly stems from the use of a cylindrical configuration, which results in a larger absorber area and hence larger reradiation loss, than the more common paraboloidal or spherical reflectors. Setting $\alpha=0.8$ and $\epsilon=0.2$ provides a slightly conservative representation of the best selective surfaces, and these values are used throughout this paper except where stated otherwise.[8-9] It may be noted that for the values chosen, collection efficiency can not exceed 0.8 pu, except possibly with a concave absorber where the cavity effect produces a slight increase in equivalent absorptivity. Evaluation of any geometry should compare actual efficiency with this 0.8 pu limit.

Absorber Shape

Other conditions being equal, the collection efficiency attainable with a flat or concave absorber is generally greater than the efficiency for a convex absorber. However, for conditions which result in high efficiency, the difference in maximum efficiency for the several shapes becomes very small. Since most practical applications would, by necessity, operate at high efficiency, the choice of absorber shape may be based on practical factors rather than efficiency. Fig. 5 shows the variation in efficiency with shape, for the same conditions as in Figs. 3 and 4.

The efficiency of a concave absorber may be either greater or less than a flat absorber depending on the relative importance of reflection loss and reradiation loss. The cavity effect for a concave absorber increases the effective values of both α and ϵ. If reflection loss is predominate, then the increased absorption results in higher efficiency for the concave absorber. If reradiation loss is predominate, then the increased emission results in lower efficiency for the concave absorber. This is illustrated by the curves in Fig. 6, where as temperature increases, the reradiation loss tends to predominate. However, it is evident that the difference in efficiency between flat and concave absorbers is small at the high efficiencies.

Reflector Diameter

The reflector diameter which results in maximum efficiency depends on the absorber shape. Typical values of the optimum ratio of diameter to focal length are shown in Fig. 5. It turned out for a wide variety of cases investigated that the optimum value of d/f for flat or nearly flat absorbers is very nearly equal to 5/3. For absorbers with relatively large values of b/a, there is more variation in the optimum value of d/f, but in no case did the optimum differ from the curve in Fig. 5 by more than 50%. Furthermore, the value of d/f becomes less critical as b/a increases, as is shown by the broad maxima of the curves in Fig. 4 for large b/a.

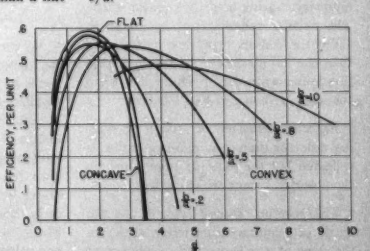


Fig. 4. Typical variation of collection efficiency versus reflector diameter

I=90 watts/sq ft
δ=0.75 degree
T=1,000 F
γ=70 degrees
α=0.8
ε=0.2

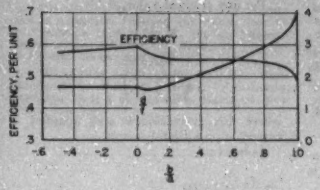


Fig. 5. Maximum collection efficiency and reflector diameter at which maximum efficiency occurs versus absorber shape ratio. See Fig. 4 for specified conditions

b/a>0, convex absorber
b/a=0, flat absorber
b/a<0, concave absorber

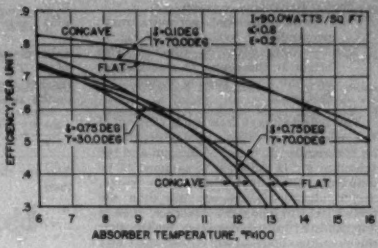


Fig. 6. Maximum collection efficiency versus absorber temperature for flat and concave absorbers

δ=tolerance on reflected rays
γ=maximum allowable angle of incidence at absorber

Absorber Temperature

Since radiation from a surface is proportional to the fourth power of temperature, the reradiation losses increase as absorber temperature increases. The resulting drop in efficiency is shown in Fig. 6.

Tolerance on Reflected Rays

The larger the tolerance, δ, the larger must be the absorber to intercept all reflected rays. Since reradiation losses are proportional to absorber area, a larger value of δ results in a smaller collection efficiency. Stated another way, the maximum absorber temperature that can be attained without dropping below a minimum acceptable efficiency depends upon the value of δ. For the conditions in Fig. 6, an efficiency of 0.6 pu can be realized at a temperature of 1,450 F if $\delta=0.1$ degree, while a temperature of only 980 F can be realized for the same conditions if $\delta=0.75$ degree.

Maximum Allowable Angle of Incidence at Absorber

For cylindrical absorbers, either convex or concave, a small value of the angle γ has the result that only the central portion of the absorber visible from any particular direction is effective for absorption. Thus for small values of γ, the total absorber area increases, and the resulting increase in reradiation losses yields lower collection efficiency. This effect is illustrated in Fig. 6 for a concave absorber.

For a flat absorber, the value of γ has no direct effect on the absorber width a. However, the reflector diameter is limited to a value such that the angle ϕ in Fig. 1 does not exceed γ. For a flat absorber, the optimum value of ϕ is very nearly equal to 45 degrees. Thus, if γ is less than 45 degrees, the reflector diameter must be less than what would otherwise be optimum. However, as shown in Fig. 6, the decrease in efficiency is small even for γ as low as 30 degrees.

Corrected Solar Radiation Intensity

Any variation in solar intensity results in a corresponding variation in the power impinging upon the absorber, yet reradiation losses remain the same. Consequently, a reduction in solar intensity causes a drop in collection efficiency. The solar intensity outside the earth's atmosphere is 126 watts/sq ft, while the intensity at the earth's surface at noon on a bright clear summer day in the United States is about 85 watts/sq ft.[18] Since the corrected value used in the analysis includes the effect of imperfect reflection, it is apparent that the value of 90 watts/sq ft used for Figs. 3 through 6 is a little low for space applications and is a little high for surface installations. However, at high efficiencies, where absorber reflection losses exceed reradiation losses, only small changes in efficiency result from moderate variations in solar intensity. For example, for the same condition as in Figs. 3 through 5, except using corrected intensities of 70 and 115 watts/sq ft, the maximum efficiencies are 0.54 and 0.64 pu, respectively.

Future Analysis

All of the foregoing results are, of course, applicable to the cylindrical configuration of Fig. 1 only. It is expected that future requirements will result in extending the analysis to other configurations, such as a paraboloidal (cup-shaped) reflector with a circular or spherical absorber. The general method of analysis will be the same as described in this paper. However, the results for other configurations may be quite different. It is hoped that the results of subsequent analyses can be published in the not too distant future.

Conclusions

The important conclusions that can be drawn for the cylindrical configuration of Fig. 1 are as follows:

1. An absorber surface having high absorptivity for solar radiation and low emissivity in the infrared region results in significantly higher collection efficiency than with a nonselective surface.

2. In most cases of practical significance, collection efficiency does not depend to any great significance upon absorber shape. However, for each absorber shape there is an optimum ratio of reflector diameter to focal length which results in maximum efficiency. For flat or nearly flat absorbers, the optimum ratio of diameter to focal length is close to 5 to 3. For the more nearly circular absorbers, the optimum ratio is generally greater than 5 to 3, is more variable with changes in other conditions, but is less critical because of a broad maximum.

3. The absorber temperature range that can be attained with reasonably high collection efficiency depends on the accuracy of manufacture and alignment of the reflector. Relatively high precision is required for absorber temperatures in the range of 1,200 to 1,500 F, while moderate errors can be tolerated below 1,000 F.

4. Moderate variations in solar intensity and in the maximum allowable angle of incidence at the absorber have relatively minor effects upon maximum attainable collection efficiencies.

Appendix. Methods of Calculation

Calculations are based on the geometry of Fig. 1, and, where applicable, assume unit length perpendicular to the plane of Fig. 1. All angles are expressed in degrees. Symbols used are those discussed under "Definition of the Problem" plus the following:

ϕ and θ = angles shown in Fig. 1
β = angular diameter of the sun, approximately 32 minutes
r = distance from an edge of the reflector to the focal point
c = projection, normal to the length r, of that portion of the absorber width which is effective in absorbing rays parallel to r

σ = Stefan-Boltzmann constant

P = net power absorbed per-unit length perpendicular to the plane of Fig. 1

x and y = rectangular co-ordinates with origin at the center of the reflector and the y-axis passing through the focal point

The reflected rays from each element of the reflector fall within a wedge depending on the angular diameter of the sun and the tolerance δ. If radiation is assumed to be uniformly distributed within each wedge, and if absorber losses are less than the power incident upon the absorber, then if any reflected radiation misses the absorber the escaping radiation will exceed the losses which would be incurred by increasing the size of the absorber. On the other hand, if the absorber is larger than necessary for the interception of all reflected rays, then the absorber losses will be larger than necessary. In either case, the collection efficiency will suffer. The optimum size absorber is just big enough to intercept all reflected rays. The critical condition which establishes the size of the absorber is the wedge of reflected rays from the edge of the reflector. This particular wedge travels farthest to reach the absorber, thus spreading out the most, and strikes the absorber at least as obliquely as any other wedge.

The equation of the parabolic section is

$$y=\frac{x^2}{4f}$$

The distance from the focal point to a point on the reflector is given by:

$$\sqrt{x^2+(f-y)^2}=f+y$$

from which

$$r=f+\frac{d^2}{16f}$$

Dividing by f to convert to dimensionless ratios yields

$$\frac{r}{f}=1+\frac{1}{16}\left(\frac{d}{f}\right)^2$$

An expression for ϕ which is convenient for execution on a digital computer is

$$\phi=90^\circ-\tan^{-1}\frac{f-\frac{d^2}{16f}}{\frac{d}{2}}$$

$$=90^\circ-\tan^{-1}\left(\frac{2f}{d}-\frac{d}{8f}\right)$$

The remaining calculations depend on the shape of the absorber.

Flat Absorber

The ratio of d to f is limited by the restriction that $\phi \leq \gamma \leq 90$ degrees. This restriction assures that the maximum angle of incidence at the absorber does not exceed γ. Next, $c=a\cos\phi$, or $c/f=a/f\cos\phi$. Since β and δ are both very small,

$$\beta+2\delta=\frac{c}{r}\text{ radians}\quad\frac{180\text{ degrees}}{\pi\text{ radians}}=\frac{180}{\pi}\frac{c}{r}$$

Solving for the ratio of absorber width to focal length:

$$\frac{a}{f}=\frac{c}{f\cos\phi}=\frac{\pi}{180}\frac{r}{f}\frac{\beta+2\delta}{\cos\phi}$$

Solar power intercepter pu focal length is Id/f. The unobstructed portion which reaches the reflector and is then reflected to the absorber is given by

$$I\frac{d-a}{f}=I\left(\frac{d}{f}-\frac{a}{f}\right)$$

The net absorbed power pu focal length is then

$$\frac{P}{f}=\alpha I\left(\frac{d}{f}-\frac{a}{f}\right)-\epsilon\sigma T^4\frac{a}{f}$$

Finally,

$$\text{Efficiency}=\frac{P/f}{Id/f}=\alpha\left(1-\frac{a}{d}\right)-\frac{\epsilon\sigma T^4}{I}\frac{a}{d}$$

Concave Absorber

The best match between absorber and reflector is obtained with $\theta=\phi+\gamma$ with the restriction that $\theta\leq 90$ degrees. Then:

$$\frac{b}{a}=\frac{1}{2}\tan\frac{\theta}{2}$$

Now

$$\beta+2\delta=\frac{180}{\pi}\frac{c}{r}$$

or

$$\frac{c}{f}=\frac{\pi}{180}\frac{r}{f}(\beta+2\delta)$$

If $\phi+\gamma\leq 90$ degrees, then

$$\frac{a}{f}=\frac{c}{f\sin\gamma}=\frac{\pi}{180}\frac{r}{f}\frac{\beta+2\delta}{\sin\gamma}$$

If $\phi+\gamma>90$ degrees, but with $\phi<90$ degrees, then

$$\frac{a}{f}=\frac{c}{f\cos\phi}=\frac{\pi}{180}\frac{r}{f}\frac{\beta+2\delta}{\cos\phi}$$

ϕ can not exceed 90 degrees. If it did, rays from the edge of the mirror could not reach the absorbing surface.

The cavity effect for the concave absorber causes the effective values of absorptivity and emissivity to increase. The following expressions are readily derived, where the effective values are denoted by primes:

$$\alpha'=\alpha\frac{\frac{\pi}{180}\frac{\theta}{\sin\theta}+\frac{90-\gamma}{90+\gamma}}{\frac{\pi}{180}\frac{\theta}{\sin\theta}\alpha+\frac{180}{90+\gamma}-\alpha}$$

$$\epsilon'=\frac{\epsilon}{\epsilon+\frac{180}{\pi}\frac{\sin\theta}{\theta}(1-\epsilon)}$$

Then

$$\frac{P}{f}=\alpha' I\left(\frac{d}{f}-\frac{a}{f}\right)-\epsilon'\sigma T^4\frac{a}{f}$$

and

$$\text{Efficiency}=\frac{P/f}{Id/f}=\alpha'\left(1-\frac{a}{d}\right)-\frac{\epsilon'\sigma T^4}{I}\frac{a}{d}$$

For any selected value of b/a,

$$\theta=90+\tan^{-1}\frac{\frac{1}{2}-\frac{b}{a}}{\sqrt{\frac{b}{a}\left(1-\frac{b}{a}\right)}}$$

where this expression is well adapted to execution on a digital computer.

As before:

$$\frac{c}{f}=\frac{\pi}{180}\frac{r}{f}(\beta+2\delta)$$

If 180 degrees $-\theta-\phi>\gamma$,

$$\frac{a}{f}=\frac{c}{f\sin\gamma}=\frac{\pi}{180}\frac{r}{f}\frac{\beta+2\delta}{\sin\gamma}$$

If 180 degrees $-\theta+\phi>\gamma>180$ degrees $-\theta-\phi$,

$$\frac{a}{f}=\frac{c}{f}\frac{2}{\sin\gamma+\sin(180^\circ-\theta-\phi)}$$

$$-\frac{\pi}{90}\frac{r}{f}\frac{\beta+2\delta}{\sin\gamma+\sin(180^\circ-\theta-\phi)}$$

If 180 degrees $-\theta+\phi<\gamma$ and $\phi+\theta-180$ degrees $<\gamma$,

$$\frac{a}{f}=\frac{c}{f}\frac{2}{\sin(180^\circ-\theta+\phi)+\sin(180^\circ-\theta-\phi)}$$

The angular sum, $\phi+\theta-180$ degrees, can not exceed γ; if it did, rays from the edge of the reflector could not reach the absorber surface at an angle of incidence less than γ. If $\theta\leq 90$ degrees,

$$\frac{P}{f}=\alpha I\left(\frac{d}{f}-\frac{a}{f}\right)-\epsilon\sigma T^4\pi\left(1-\frac{\theta}{180}\right)\frac{a}{f}$$

If $\theta>90$ degrees,

$$\frac{P}{f}=\alpha I\left(\frac{d}{f}-\frac{a}{f}\sin\theta\right)-\epsilon\sigma T^4\pi\left(1-\frac{\theta}{180}\right)\frac{a}{f}$$

Finally,

$$\text{Efficiency}=\frac{P/f}{Id/f}=\frac{P}{Id}$$

Nonselective Absorber Surface

For the special case where $\alpha=\epsilon$, it is apparent for all absorber shapes that the various expressions for net absorbed power all reduce to the form $P/f=\epsilon K$, where K is independent of ϵ. It is thus obvious that the greatest net absorption will occur if $\alpha=\epsilon=1.0$; i.e., a black absorbing surface is better than any other nonselective surface.

References

1. Solar-Powered Thermoelectric Generator Design Considerations, N. F. Schuh, R. J. Tallent. *AIEE Transactions*, see pp. 345–52 of this issue.

2. Notes on Optical Design Principles, Concentration Ratios, and Maximum Temperature of Parabolic Solar Furnaces, R. W. Bliss, Jr. *Technical Memorandum HDW-TM-56-1*, Holloman Air Development Center, N. Mex., July 1956.

3. The Flux Through the Focal Spot of a Solar Furnace, P. D. Jose. *Technical Memorandum HDRRM-TM-57-3*, Holloman Air Development center, Apr. 1957.

4. Calculation of the Concentration of the Solar Radiation Through the Fucal Spot of a Paraboloic Mirror, A. W. Simon. *Technical Memorandum HDGR-57-10*, Holloman Air Development Center, Aug. 1957.

5. Utilization of Solar Furnaces in High Temperature Research, P. Duwez. *Transactions*, American Society of Mechanical Engineers, New York, July 1957.

6. Selective Radiation, H. Tabor. *Bulletin*, Research Council of Israel, Jerusalem, Israel, vol. 5, sect. A, 1955.

7. Solar Energy Design, H. Tabor. *Ibid.*, vol. 5, sect. C, 1955.

8. The Preparation of a Selectively Black Surface for Use in Collecting Solar Energy, T. A. Unger. *Doctoral Thesis*, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., 1955.

9. Filmed Surfaces for Reflecting Optics, O. Hass. *Journal*, Optical Society of America, New York, N. Y., Nov. 1955.

10. Handbook of Geophysics for Air Force Designers (book). Air Force Cambridge Research Center, Cambridge, Mass., 1957, chap. 16.

A Practical Standard Transistorized Optimum Response Controller

KAN CHEN
ASSOCIATE MEMBER AIEE

D. R. LITTLE
ASSOCIATE MEMBER AIEE

Synopsis: The principle of nonlinear optimum response control has found a new and unique application in the development of a standard controller. The transistorized controller can work in a great variety of feedback systems because it has negligible time delay. Optimum response to a specified step input can be obtained very easily by experiment because linear switching is used. The response is still nearly optimum when the input is not of the specified form and magnitude, or when the system parameters change with environmental variations. Solution of practical problems in the development are discussed in the paper. A mathematical proof is given to show that optimum response to a specified step input, for the systems under consideration, can always be obtained with linear switching.

THE PRESENT METHODS of designing practical control systems leave much to be desired insofar as engineering time is concerned. In most cases, the compensation networks and at least one of the control amplifiers are custom-made to meet individual system specifications. Although systems designed in this manner give fairly satisfactory results, two questions are ever present:

1. Is this the best system which can be produced for the material cost involved, i.e., is there maximum utilization of all components?

2. Is there a possibility of producing one controller, amplifier and compensation included in one package, which will be applicable in a wide range of systems instead of designing a special controller for each system?

Hitherto, no controller appears to have been designed which will answer both questions. It has been proved analytically that for second order systems, i.e., systems of two time constants, the best transient response for a given available forcing can be obtained by using an ideal on-off amplifier with a nonlinear function generator, the so-called bang-bang control.[1-4] In other words, this leads to minimization of necessary forcing to meet certain specifications on recovery time, and, in general, minimum forcing means minimum material cost for the controller. Thus, this would seem to be an answer to the first question above. However, the use of a nonlinear function generator is objectionable. Moreover, an ideal on-off amplifier remains to be developed. As to the second question, a number of attempts have been made to develop some form of a standard controller which is more than just a standard amplifier, to be used in a great variety of control systems. Of these, most, if not all, have been designed from the standpoint of linear theory and, although some have succeeded in insuring adequate control under small disturbances,[5,6] none have yet been developed which can provide nearly optimum control under large disturbances. Furthermore, the adjustments of compensation networks in these controllers are complicated and require the engineering judgment or calculation of experienced personnel.

The new controller described in this paper is the result of an attempt to realize a practical controller which would answer both of the questions mentioned above. Since the on-off type of operation produces the best transient response, it was decided to devise a practical on-off type of standard controller without resorting to complicated circuitry. This then would guarantee the realization of maximum utilization of components, at least under the condition of worst system disturbance when maximum utilization of components is most desired. It was also determined that the controller should be able to control many different systems with relatively few adjustments. The controller was to be a standard controller in the strictest sense; that is, even an inexperienced technician can obtain the best system performance with the few adjustments provided by the controller. Naturally, such a controller has many advantages. Not only is the mass production of such controllers and the maximum utilization of components feasible, but the design engineering and construction time can be reduced to a bare minimum.

Optimum Response

Let us review briefly the idea of nonlinear optimum response control before we extend it to the development of a standard controller. Fig. 1 shows a nonlinear optimum response system with a second-order overdamped controlled element, the type most commonly encountered in industrial applications. The positive and negative forcings of the

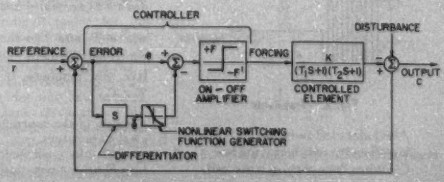


Fig. 1. On-off control system

Paper 59-838, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 10, 1959; made available for printing May 13, 1959.

Kan Chen is with the Westinghouse Electric Corporation, East Pittsburgh, Pa., and D. R. Little is with the Curtiss-Wright Corporation, Quehanna, Pa.

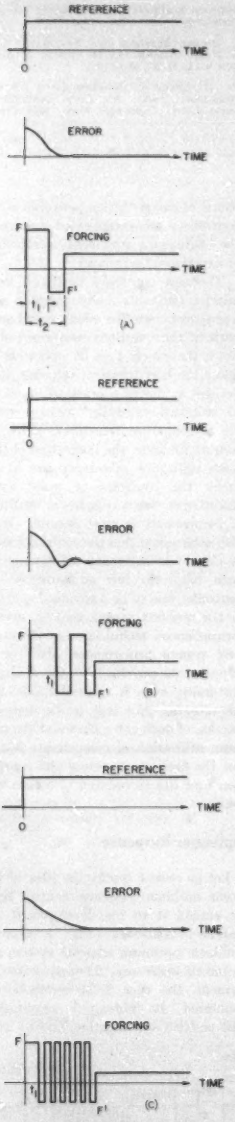


Fig. 2. Transients

A—Under ideal switching
B—When first switching is too late
C—When first switching is too early

on-off amplifier are unsymmetrical in the general case. Thus, the case of a single-ended amplifier is definitely included. In this control system, if a proper non-linear switching curve is used to determine when to switch from positive to negative forcing, then the graph of the reference step input, the error, and the forcing with respect to time is shown in Fig. 2(A). These curves will then describe the so-called "optimum" case where t_1 denotes the time at which forcing was switched from full positive to full negative and t_2 when forcing is switched to a value just sufficient to maintain the desired output. In the controller scheme shown, the actual system will jitter at a very high frequency under steady state, and the durations of positive and negative controller output are so proportioned that zero error is maintained. It has been shown mathematically that the optimum case described corresponds indeed to minimum time for the system error to reach and remain zero subject to the constraint of the given maximum available forcing of the controller.[3] If the switching time, t_1, is too late, then the curves of Fig. 2(B) describe the behavior of the system. When the switching time, t_1, is too early the curves of Fig. 2(C) characterize the system behavior. In either case, the response time is longer than that in the optimum case shown in Fig. 2(A).

A frequently used example to illustrate the optimum response principle is that the quickest way to race an automobile from one stop light to another is to press down the accelerator to the floor and hold it there for a critical period of time and then to apply full braking. In this manner, an automobile can beat another with a more powerful engine but not using full power and braking during the race.

Mathematical Description

The performance of the optimum system in Fig. 1 can be described quantitatively in a phase plane plot of the first derivative of the error $\dot{e}$ versus the error e. From Fig. 1 we can derive the following expression:

$$T_1T_2\,\ddot{c}+(T_1+T_2)\,\dot{c}+c=Kf$$

where T_1 and T_2 = time constants of the controlled element, assumed to be positive real numbers, or overdamped, and $T_1<T_2$.

K = gain of the controlled element
c = system output
$f\begin{cases}=F>0 \text{ for positive forcing}\\=F'<0 \text{ for negative forcing}\end{cases}$

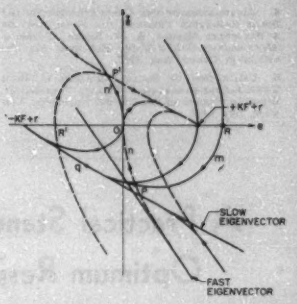


Fig. 3. Phase-plane trajectories of the ideal on-off system

Referring again to the closed loop system of Fig. 1 and assuming that the reference input r is a step function (i.e., r is a constant during the system transient), then

$$e=r-c$$
$$\dot{e}=-\dot{c}$$
$$\ddot{e}=-\ddot{c}$$

so that

$$T_1T_2\ddot{e}+(T_1+T_2)\dot{e}+e=K_1 \qquad (1)$$

where

$$K_1=-Kf+r$$

The general solution of equation 1 is

$$e=Ae^{-t/T_1}+Be^{-t/T_2}+K_1 \qquad (2)$$

where the constants A and B depend on the initial conditions and e is base of natural logarithms.

We then find the derivative of the error to be

$$\dot{e}=\frac{A}{T_1}e^{-t/T_1}-\frac{B}{T_2}e^{-t/T_2} \qquad (3)$$

Eliminating t from equations 2 and 3 yields

$$T_1\ln\frac{(e-K_1)+T_2\dot{e}}{A(1-T_2/T_1)}=T_2\ln\frac{(e-K_1)+T_1\dot{e}}{B(1-T_1/T_2)} \qquad (4)$$

which describes the trajectories in the phase plane. A few trajectories, each corresponding to a different set of values of A and B are shown in Fig. 3. The solid curves are indicative of positive forcing and the dash curves of negative forcing. The arrows indicate the direction taken by the operating point with increasing time.

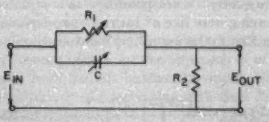
It is important to point out that the trajectories near the nodes $(K_1, 0)$ approach asymptotically to the slow eigenvector with an absolute slope equal to

$1/T_2$, and those far away from the nodes diverge asymptotically from the fast eigenvector with an absolute slope equal to $1/T_1$.

Switching Curve

Suppose a step input R is initiated in the reference r such that $e=R$ and $\dot{e}=0$ initially. If, as the result of this condition, positive forcing is initiated, the operating point will move along curve m and, unless switching occurs at point P, the operating point will move towards and eventually reach the point $(-KF+r, 0)$ along path q. However, if at point P the system switches from full positive forcing to full negative forcing, the operating point will follow path n into the origin. The system will come to rest at the origin; that is, the error and the derivative of error are zero simultaneously if from this time on, the average output of the controller is just enough to maintain the output at its desired level. If the system is switched at any point other than point P, the operating point will not be able to reach the origin without switching at least once more since the curve n is the only negative forcing curve passing through the origin. If the initial step is negative and places the operating point at R', then a similar case exists where P' is the switching point and n' is the curve passing through the origin. Here the negative forcing period precedes the positive forcing period.

Since a wide range of positive step disturbances are possible and since it is desirable to come into the origin via the curve n, it is apparent that the switching from positive to negative forcing must occur when the path of the operating point intersects curve n. Similarly, for the case of a negative step the curve n' is the path that should be followed into the origin. Hence, the combined curves nn' should be the so-called "switching curve." This means that when the operating point is to the right of the switching curve, positive forcing should be employed, and to the left negative forcing should be employed in order to accomplish optimum response in all cases. One way to achieve the switching curve is indicated in Fig. 1 by the blocks between the error and the ideal on-off amplifier.



$$\frac{E_{OUT}(S)}{E_{IN}(S)} = \frac{R_2}{R_1+R_2} \cdot \frac{1+R_1CS}{1+\frac{R_1R_2}{R_1+R_2}CS}$$

Fig. 4. Derivative network

Solving the Practical Difficulties

The optimum response system described has attracted the interest of many engineers and scientists in the field of advanced control development, as is evidenced by the large number of recent technical papers on this subject. However, the system has not been used extensively in practice so far. One may summarize the doubts about the practicalities of such a system as follows:

1. The system response is optimum for only step and ramp inputs or disturbances.

2. The controlled element may not be linear, second order, and overdamped, as assumed in the foregoing analysis.

3. Pure differentiation of error assumed in the analysis is difficult to achieve and tends to accentuate noise.

4. The high-frequency jittering of the on-off amplifier may cause intolerable ripple in the output signal.

5. The ideal on-off amplifier may be difficult to approach with practical circuits.

6. The nonlinear function generator, shown in Fig. 1, used in such a system is objectionable from the standpoints of cost and adjustability.

7. The optimum response may be critically affected by changes of system parameters due to load or environmental variations.

These doubts are quite reasonable for one who has had extensive practical experience in developing control systems. The remaining part of this report will explain how the practical difficulties are overcome or circumvented. Here are the point-to-point answers to the objections listed above:

1. The worst input and disturbance to many control systems are actually in the form of a step or ramp. In fact, most industrial control systems are specified in terms of such inputs.

2. Most industrial control systems are higher than second order mainly because of the appreciable time delay attributed to the power amplifiers and the compensation networks. If the ideal on-off amplifier can provide sufficient control power for the output element without any significant time delay and no compensation other than that in the standard configuration given in Fig. 1 is used, most system controlled elements are essentially second order and overdamped. For example, the time constants of the exciter and of the generator are the essential delays in a voltage regulating system, and the time constants of the generator and of the motor are the essential delays in a speed control system. As to the effects of slight nonlinearities in the controlled elements, they are best determined by an analog computer.

3. Of course, pure differentiation is not feasible to achieve in practice. However, it can certainly be approximated by a simple resistance-capacitance (RC) network, such as given in Fig. 4. The small time lag in this derivative network as well as the large time delays in the controlled element should tend to alleviate the problem of high-frequency noise.

4. The output elements, because of their inherent time delays, should serve as low-pass filters to attenuate the jittering in the on-off amplifier output under steady state. As long as the jittering frequency is high enough, the ripple in the system output signal should not be discernible. This is true in the same sense that periodic output waveforms of such devices as magnetic amplifiers and thyratrons do not cause discernible ripple in the system output. Furthermore, the jittering of the on-off amplifier can be avoided if the on-off amplifier is allowed to have a narrow linear range and provided that the system is stable within this linear range.

5. Most on-off amplifiers which have been considered in the literature to achieve optimum system response are electromechanical relays. Their wear and tear, appreciable switching time, and hysteresis characteristics make their use in control systems very undesirable. The problem can be circumvented if transistors are used as the basic components of the amplifier. The requirements of the on-off amplifier will be discussed later in a separate section.

6, 7. These problems are dealt with in the next section.

Linear Switching

To avoid the cost and the inflexibility of a nonlinear function generator, we shall consider replacing the nonlinear curve in Figs. 1 and 3 by a straight line. If the curve nn' in Fig. 3 is substituted by a straight line AA' going through the points O and P as shown in Fig. 5, the

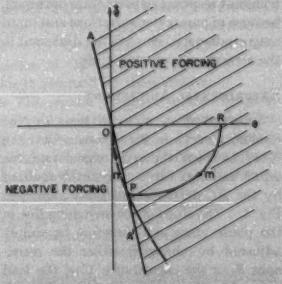


Fig. 5. Phase-plane trajectories of the linear on-off system

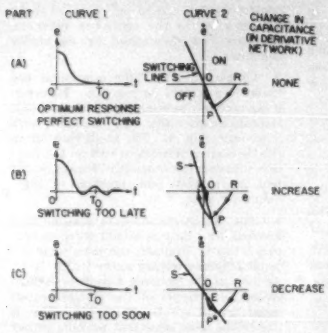


Fig. 6. Optimum response adjustments

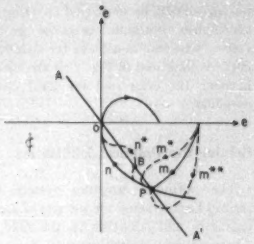


Fig. 7. Apparently possible trajectories crossing the optimum switching line

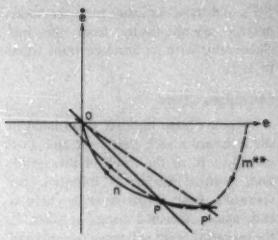


Fig. 8. More than one intersections of m and n trajectories in the fourth quadrant

curve m still lies completely in the positive forcing region and the curve n still lies completely in the negative forcing region, as they did in Fig. 3. Hence, the system still responds in an optimum manner to an input step change exactly equal to R. Suppose R is the worst step input or disturbance for which the system specification is written, then to a less severe input or disturbance the response will not be optimum but is likely to be within the specified recovery time for the worst case. The observation that the ideal switching curve n approaches the fast eigenvector, a straight line, in the region remote from the origin of the phase plane leads one to suspect that replacement of n by an appropriate straight line would result in nearly optimum or quasi-optimum response for all step inputs. Furthermore, with system parameters changing with load and environmental conditions, it is senseless to attempt achieving exact optimum response under all conditions in a practical application. How much the system transient response departs from optimum, because of parameter variations and input magnitude variations will be discussed in the section on computer study.

Experimental Adjustment

Substituting a straight line for the nonlinear curve in Fig. 1 means that the block between the error signal and the on-off amplifier can be physically realized by the RC derivative network shown in Fig. 4. The slope of the switching line in the phase plane can therefore be simply adjusted by changing either the resistance R_1 or the capacitance C of the lead network. However, changing R_1 will change the ratio of the lead and lag time constants as well as the gain of the derivative network. Therefore, it is advisable to adjust the slope of the linear switching line by changing the capacitance alone.

If the controlled element is second-order and overdamped, the only parameter on the controller to be adjusted is the capacitance alone as far as adjustment for optimum transient response is concerned. Of course, the position of this "simple knob" on the controller can be determined by calculation based on the gain and time constants of the controlled element and the positive and negative forcing of the controller. However, a more practical way of adjusting this single knob is by experiment. The procedure consists of feeding the largest expected, or specified, step input or disturbance to the system and examining the time response. If there is an overshoot, the time of the first switching should be made earlier by increasing the capacitor C. If the response has a long tail, the time of the first switching should be made later by decreasing the capacitor C. The optimum response corresponds to no overshoot nor tail, that is, dead-beat response. The transient responses of these 3 cases are illustrated in Fig. 6, together with their corresponding phase-plane trajectories. The procedure is so simple that even an inexperienced technician can determine the correct setting of the capacitor to obtain optimum response. Another experimental method of determining the correct setting of the capacitor is to see that one and only one switching of the controller output occurs between the beginning and the end of the transient.

Part C of Fig. 6 shows that the "end-point" phenomenon[2] occurs when the capacitance of the derivative network is too large. After the end-point E is reached the controller output oscillates at a very high frequency and with such an on-off time ratio that the operating points proceeds along the switching line S to the origin of the phase plane. The response for this part of the transient is an exponential function of time with a time constant equal to R_1C of the derivative network. Although it takes theoretically infinite time to reach steady state, the time it takes to reach and remain with a specified band around the final value is finite.

A similar statement may be made in regard to part B of Fig. 6, the overshoot case.

Mathematical Proof

Consider an on-off system with a second-order, overdamped controlled element. Although linear switching will not result in optimum response for all magnitudes of step inputs, it can be proved mathematically that for a given step input, an on-off controller employing a proper linear switching function can always achieve optimum response. The previous discussion, based on the phase-plane trajectory in Fig. 5, has illustrated the possibility of obtaining optimum response with a linear switching line. However, this is possible only if the straight line AA' linking the origin, O, and the optimum switching point, P (Fig. 5), puts the trajectories m and n entirely in the two separate half-planes divided by AA'. For example, if the trajectory corresponding to negative forcing were like n^* instead of n, depicted in Fig. 7, the use of AA' as the switching

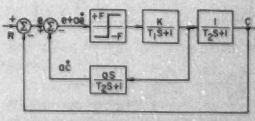
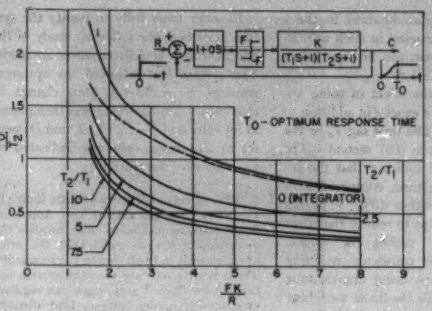


Fig. 9. On-off system using exact ė



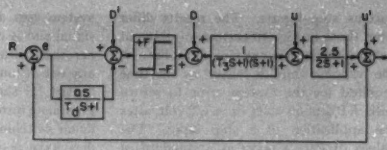


Fig. 10 (left). Optimum response time

Fig. 11 (above). On-off system using approximate $\dot{e}$

line would cause the on-off controller to switch back to positive forcing at B, whereas negative forcing is required to give optimum response. For the same reason, if the trajectory corresponding to positive forcing were like m^* or m^{**} instead of m, as depicted in Fig. 7, it would not be possible with linear switching to obtain optimum response to a given step input. The purpose of the following mathematical proof is to show that none of the trajectories m^*, m^{**}, and n^* can occur.

1. If the trajectory n^* occurred, the rate of change of the slope $d\dot{e}/de$ with respect to time t would have to change sign. From equations 2 and 3,

$$\frac{d\dot{e}}{de}=\frac{d\dot{e}}{dt}\Big/\frac{de}{dt}=\frac{d\dot{e}}{dt}\Big/\dot{e}$$

$$=\left(\frac{A}{T_1^2}\,\epsilon^{-t/T_1}+\frac{B}{T_2^2}\,\epsilon^{-t/T_2}\right)\Big/\left(-\frac{A}{T_1}\,\epsilon^{-t/T_1}-\frac{B}{T_2}\,\epsilon^{-t/T_2}\right) \quad (5)$$

Therefore,

$$d\left(\frac{d\dot{e}}{de}\right)/dt=\frac{AB(T_1+T_2)^2}{T_1^2T_2^2\left(\frac{A}{T_1}\,\epsilon^{-t/T_1}+\frac{B}{T_2}\,\epsilon^{-t/T_2}\right)}\times\epsilon^{-\left(\frac{t}{T_1}+\frac{t}{T_2}\right)} \quad (6)$$

This expression does not change sign as a function of t; and hence the trajectory n^* cannot occur.

2. It is obvious that the trajectory m^* cannot occur: for the approach of m^* to P from the left implies the impossible requirement of $\dot{e}$ increasing with time for $\dot{e}$ being negative.

3. The occurrence of m^{**} would imply a relative configuration of m^{**} and n as shown in Fig. 8. This seems possible because both m^{**} and n are concave upward and intersect the e-axis at the right points. These trajectories now have three intersections in the fourth quadrant of the phase plane. If this should occur, P would still be the optimum switching point and the switching line OP would still put m^{**} and n in two separate half planes. However, the use of another switching line OP' would result in a dead-beat response with one switching during the transient. This would mean that the dead-beat transient response would not correspond uniquely to the optimum response.

It is recalled that the phase-plane trajectories for the second-order systems under consideration always diverge asymptotically from the fast eigenvector and approach asymptotically to the slow eigenvector. Therefore, in the fourth quadrant of the phase plane, the m trajectories should have a slope between $+\infty$ and $-1/T_2$ and the n trajectories should have a slope between $-1/T_1$ and $-\infty$. Since $-1/T_2$ is larger than $-1/T_1$ algebraically, the slope of m should always be larger than that of n in the fourth quadrant. This condition is violated at the intersection point P in Fig. 8 where the slope of m^{**} is less than that of n. Therefore m^{**} cannot occur. The same reason rules out all possibilities but that of only one intersection of the m and n trajectories in the fourth quadrant.

Thus, it has been shown mathematically that optimum response can always be achieved by using proper linear switching and that optimum response corresponds uniquely to dead-beat response for the systems under consideration.

Computer Study

The objective of the computer study was twofold: 1. To determine the optimum system response time as a function of the available forcing and of the parameters of the controlled element when a system exactly like that in Fig. 1 is obtainable; 2. To determine the effects upon the optimum response time in a typical case when the exact control system of Fig. 1 has to be modified for various practical reasons given in the previous sections. The analog computer was used because analytical methods were time-consuming and the 1 to 2% error in the computer results is considered within reasonable tolerance for practical purposes.

It has been pointed out in the previous section that the optimum response provided by the system in Fig. 1 can be exactly reproduced through proper linear switching if the input step is fixed. Hence, the first objective of the computer study was attained by conducting computer experiments on the system shown in Fig. 9. Here the error derivative $\dot{e}$ ($=-\dot{c}$) is exact since no disturbance is introduced into the output element. The system is optimized by adjusting a, which controls the slope of the switching line, until the system error comes to rest after exactly one switching of the on-off amplifier. The realization of optimum response can be checked further by seeing that the output transient has neither overshoot nor a tail. The information on the exact optimum response time is presented in the form of a family of curves in Fig. 10. The parameters are normalized in order to maximize the range of application of the curves. Symmetrical forcing is assumed for these results.

The second objective of the computer study was attained by performing computer experiments on a typical system with ± 50 volts of available forcing and controlled element with the following nominal parameter values:

Gain $K=2.5$
Time constants $T_1=1$ second
$T_2=2$ seconds

The results are tabulated in the Appendix.

Part I of the Appendix shows the response time of the optimized system for

various step inputs. The results differ from those in Fig. 10 because of different definitions of response time. In Fig. 10, the response time is defined as the time required for the system error to return with $\pm 1\%$ of 50 volts, or $\pm 1/2$ volt, after the application of a step input. The values in Part I serve as a standard of comparison for the following results.

Part II shows the effect of variation of step inputs when the linear switching line is adjusted only for optimum response following a step input of 50 volts and is held fixed thereafter for any other inputs. The responses of the system to step inputs other than 50 volts are therefore non-optimum. However, there is little difference between the response times of Part I and their corresponding ones in Part II. This confirms the belief that the exact switching curve nn' of Fig. 3 is nearly linear for a second-order, overdamped, controlled element, and hence nn' can be approximated quite satisfactorily by a straight line. The results in Part II are significant in that they justify the use of linear switching. Note that the method of obtaining the exact $\dot{e}$ in Fig. 9 is good only when there is no load disturbance on the controlled element. Therefore, this method cannot be used in practice. Fig. 11 shows the basic system simulated on the computer for all of the remaining parts of the computer study. This system uses an approximation to the $\dot{e}$ signal and is constructed in such a manner that disturbances can be injected in any one of four locations.

Part III shows the effect of using approximate instead of exact $\dot{e}$ in addition to linear switching. The time lag T_d of the derivative network is 0.01 second. The results of this part indicate that the increases of response time, compared with Part I, are well within 10%. The lag effect due to T_d helps to reduce the effect of high-frequency pickup or noises. Fig. 12 shows the transients resulting from a 10-volt step, using approximate $\dot{e}$ in the system, while the linear switching line has been adjusted to give optimum response for a 50-volt step input.

Parts IV and V show the effect of an appreciable third time constant T_3. In Part IV, $T_3 = 0.1$ second and in Part V, $T_3 = 0.5$ second. Results again appear in the Appendix. These results indicate that a third time constant (T_3) of 1/10 the smaller time constant, $1/10T_1$, does not appreciably affect the system response; however, for $T_3 = 1/2T_1$, the system is adversely affected. In such cases, the system should be treated as essentially third order. Systems higher than the second order can be optimized by measuring the second and higher derivatives of the error. However, they will not be considered in this paper.

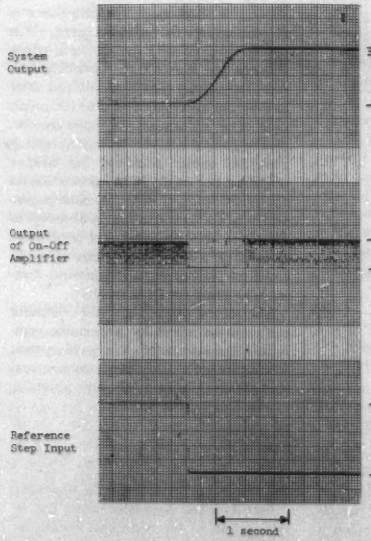


Fig. 12. Computer result, 10-volt step input with approximate $\dot{e}$, system optimized for 50-volt step input

In order more fully to justify the system described, the following study of the effects of drift and disturbances are made. The system was adjusted for optimum response to a 50-volt step change in reference. The reference was then placed at 50 volts and T_3, U, U', D, and D' were set to zero unless otherwise indicated by the following description.

1. The effects of disturbances are shown in Parts VI through IX of the Appendix. Some response times are zero because the disturbances are so small that the error never exceeds ± 1 volt during the transients. These results indicate that the system optimized for 50-volt step input can regulate satisfactorily under various load disturbances.

2. The effects of drift and 60-cycle pickup upon the steady-state error are shown in Parts X through XIII. Note that this experiment involved adding a d-c level at points indicated in Fig. 11. This did not involve step or ramp disturbances.

These results indicate that the optimum system has good accuracy, as may be expected due to the extremely high gain of the on-off amplifier. The drift at the input end of the controller is equivalent to error in the system measurement device and is therefore directly reflected in the system error. The ability of the system to regulate satisfactorily under a large amount of 60-cps (cycles per second) pickup is of interest. It shows that the lead network does not accentuate high-frequency noises to an unacceptable extent.

The last portion of the computer study, Part XIV, consisted of varying the values of T_1, T_2, and K and recording the system response to a 50-volt step change in reference. In this part, as in the parts immediately preceding, the system was set for optimum response following a 50-volt step change in reference with $T_1 = 1$ second, $T_2 = 2$ seconds, and $K = 2.5$. The response time for each condition, i.e., for various values of T_1, T_2, and K, is indicated in Part XIV of the Appendix. The actual curves for cases 9 and 10 in this part are shown in Figs. 13(A) and (B), respectively. These correspond to a change in both time constants of 30% and a change in K by 69%, representing the variation in these parameters due to the change in the resistance of copper wire in a 2-stage electrical amplifier (magnetic amplifiers or rotating machines) when the temperature is varied from −60 F (degrees Fahrenheit) to 180 F (−50 C [degrees centigrade] to 75 C). Note that although the responses in the last part of the computer study were not optimum, the response time was still well within 10% of the optimum values, which were

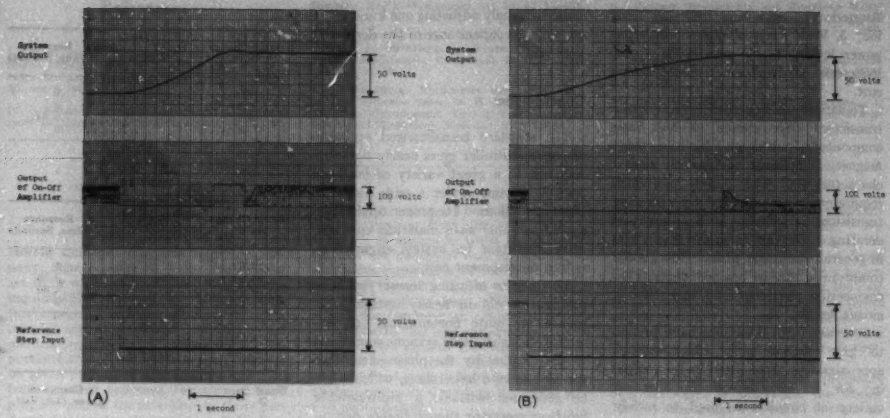


Fig. 13. Computer result, 50-volt step input, approximate $\dot{e}_i$ system optimized for 50-volt step input under normal temperature

A—System under simulated high temperature
B—System under simulated low temperature

obtained by readjusting the switching line. This shows that the practical systems under study have "flat" optimum responses; that is, their optimum responses are not critical to parameter variations.

A brief study has also been made on the computer as to the effect of slight nonlinearities, especially the saturation type, in the controlled elements. The conclusion is that the switching line can always be readjusted to give optimum response. No special adverse effects have been observed that can be ascribed to the presence of slight nonlinearities in the output elements.

In summary, the results of the computer study are quite encouraging, as evidenced by the values in the Appendix. The experience gained during the computer study confirmed the earlier prediction that the optimization of system response can be quickly achieved by simple adjustments of the switching line.

Practical Circuits

Because of its simplicity and flexibility, the lead network shown in Fig. 4 will be used as the inherent compensation of the controller. The main problem is then the physical realization of the on-off amplifier. Ideally the amplifier should have the following characteristics: 1. No delay in responding to command; 2. No delay in switching from one state of output to the other; 3.

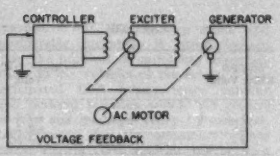


Fig. 14. Voltage regulator

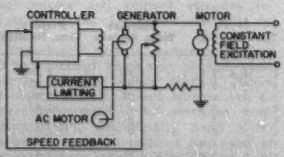


Fig. 15. Ward Leonard speed regulator

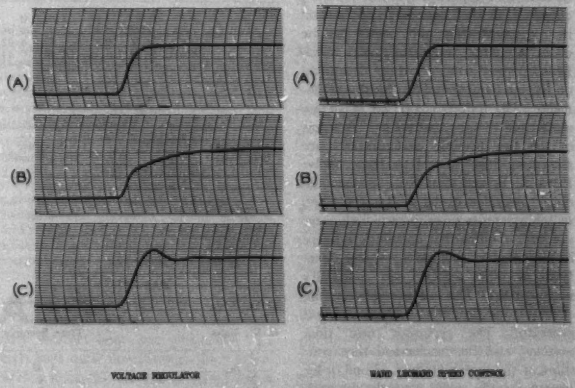


Fig. 16. Demonstration unit

Rugged; no wear and tear; 4. Long life; 5. Wide range of input and output power levels; 6. No zero drift; 7. No hysteresis or dead zone; 8. Minimum cost.

To satisfy the first three requirements, transistors were chosen as the basic components, although high-frequency magnetic amplifiers can provide similar characteristics. The fourth and fifth requirements are achieved by operating transistors as switches with appropriate derating. The last three conditions are, in general, not compatible. How good a compromise can be reached depends on engineering judgment and ingenuity of circuit design.

Several practical circuits have proved to give similar gross characteristics approximating those listed in the foregoing for an ideal on-off amplifier. A variety of circuits are used to satisfy different needs of output forms, e.g., single-ended and push-pull, d-c and a-c, single-phase and multiphase. The details and analyses of these circuits will be reported elsewhere. However, it is of interest to point out that all these circuits are functionally represented by the single block for the on-off amplifier in Fig. 1. Therefore, the approach is that of forcing a physical system to take the desirable functional form depicted by Fig. 1, with the exception of using linear switching. The difficulty in analyzing an on-off system with dead time delay and/or hysteresis is thus avoided.

Practical Demonstration

The new controller has been employed in a variety of control systems. While its industrial application will be reported elsewhere, we shall describe its application in a demonstration unit, which can be either a voltage regulator or a Ward-Leonard speed control system, depending on the mechanical and electrical arrangement. The unit is composed of an a-c motor and two d-c machines. Diagrams of the two possible configurations are shown in Figs. 14 and 15, respectively. Parts A, B, and C of Fig. 16 show the output responses of the two systems when switching occurs properly, too early, and too late. Notice that optimum response in part A gives minimum response time. Each of these systems are essentially second order. The voltage regulator time constants are the fields of the exciter and generator. The generator field time constant and the mechanical time constant of the motor produce second order system for the speed control. In both cases the system response is made optimum by simply adjusting one knob which controls the capacitance of the derivative network.

Conclusions

The standard transistorized optimum response controller gives nearly optimum response to a great variety of feedback control systems under a wide range of input conditions. Its proper application should not only save materials cost but, more important for system engineering, reduce development cost.

The results of using linear switching function to obtain nearly optimum response are quite encouraging. Its extension to higher order systems would lead to the division by the phase space into two halves by a hyperplane, which could be generated without a multivariable function generator though higher derivatives would still have to be measured.

The successful development of the standard transistorized optimum response controller has proven that it is more rewarding practically to force a physical system to take a desirable functional form by hardware synthesis rather than to analyze an on-off servomechanism including all the imperfections of the on-off device.

Appendix. Results of the Computer Study

Step Change in Reference, Volts	Seconds				
	Part I	Part II	Part III	Part IV	Part V
50	2.0	2.0	2.04	2.04	8
100	4.44	4.44	4.54	4.6	8.06
10	0.66	0.66	0.7	0.82	4.02
1	0.13	0.14	0.15	0.28	0.44

PART VI

Step Change in U, Volts	Response Time, Seconds
5	0
10	0.76
30	3.04
1	0

PART VII

Step Change in U′, Volts	Response Time, Seconds
12.5	1.02
25	1.4
75	2.5
1	0.1

PART VIII

Ramp Change in U	Response Time, Seconds
50 t	0
10 t	0
30 t	5.04
t	0

PART IX

Ramp Change in U′	Response Time, Seconds
12.5 t	0
25 t	0.78
75 t	3.36
t	0

PART X

Signal on D	Change in Error, ($\Delta\epsilon$), Volt
10	0.15
1	0.02

In Parts X through XIII, $0.5\ V = 1\%$

PART XI

Signal on D′, Volts	Change in Error, ($\Delta\epsilon_r$) Volts
10	10
1	1

PART XII

60-CPS RMS, Volts	Error, D-C, Volts
0.15	−0.04
0.3	0
0.4	+0.1
0.45	+0.2
0.52	+0.3

PART XIII

60-CPS RMS, Volts	Error, D-C, Volts
7	+0.3
4	+0.2
3	+0.05

PART XIV

T_1	T_2	K	Actual Response Time, Seconds	Ideal Response Time, Seconds
1	2.4	2.5	2.35	2.25
1	1.6	2.5	2.0	1.85
1.2	.2	2.5	2.35	2.35
0.8	.2	2.5	2.0	1.82
1	.2	.3	1.9	1.8
1	.2	.2	2.6	2.6
1.2	2.4	3.6	2.1	2.0
1.2	2.4	1.6	3.8	3.8
0.7	1.4	1.225	3.3	3.3
1.3	2.6	4.125	2.3	2.1

References

1. Nonlinear Techniques for Improving Servo Performance, D. C. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 400–31.

2. A Phase-Plane Approach to the Compensation of Saturating Servomechanisms, A. M. Hopkin. *AIEE Transactions*, vol. 70, pt. 1, 1951, pp. 631–39.

3. On the Bang-Bang Control Problem, R. Bellman, I. Glicksberg, O. Gross. *Quarterly of Applied Mathematics*, Providence, R. I., vol. 14, 1956, pp. 11–18.

4. Differential Equations with a Discontinuous Forcing Term, D. W. Bushaw. *Report no. 469*, "Experimental Towing Tank," Stevens Institute of Technology, Hoboken, N. J., 1953.

5. Process Instruments and Controls Handbook (book), D. M. Considine. McGraw-Hill Book Company, Inc., New York, N. Y., 1957, section 5.

6. On the Synthesis of Optimalized Control Loops, C. Kessler. *Regelungstechnik*, Dec. 1954–Feb. 1955.

7. Discontinuous Automatic Control, I. Flügge-Lotz. Princeton University Press, Princeton, N. J., 1953.

Discussion

Harold Chestnut (General Electric Company, Schenectady, N. Y.): The authors are to be complimented on a most interesting and valuable paper, in which the practical aspects of performance that can be obtainable with standardized controller equipment are emphasized. In stressing the importance of cost and reliability as factors to be given proper consideration by the controller system designers, the authors have properly placed the attention on those portions of the controller system design which in the past have not received proper attention, namely, reliability and cost. Through the use of standardized components in controllers, it is to be hoped that improvements in both reliability and cost can be made which far outweigh the minor degradation in performance of the order of 10% which appears to have been obtained in the results stated for this paper.

In their paper, the authors have answered many questions concerning the control of second order systems. The writer would appreciate hearing from the authors what approach is suggested when higher order systems are employed and in systems in which the parameters of the system may change of the order of plus or minus 20%. Have the authors been able to develop criteria for obtaining an equivalent second order system, for systems actually having higher orders up into the fourth and fifth order category which equivalent second-order systems prove to have the satisfactory characteristics for describing the actual system performance?

Kan Chen and **D. R. Little**: The authors wish to thank Mr. Chestnut for his comments on the paper. It is heartening to hear an experienced control expert agreeing with the authors on the theme of this paper, which stresses the practical aspects of nonlinear control systems.

Part XIV of the Appendix gives the effects of parameter changes in the order of plus and minus 30% for a second order system controlled by the standard transitorized optimum response controller. For higher order systems, the extension of the linear switching approach means to divide the phase space into two halves by a hyperplane. The authors have studied the case of an essentially third order system on an analog computer. It was not difficult to obtain the exact $\dot{\epsilon}$ and $\ddot{\epsilon}$ (first and second derivatives of the error signal) on a computer.

A converging experimental procedure was developed to determine the proper linear combination of ϵ, $\dot{\epsilon}$ and $\ddot{\epsilon}$ in the compensation network in order to achieve optimum response to a step input. This means the determination of the proper switching plane in the three dimensional phase space by experiment. However, in a practical circuit, it was difficult to obtain both $\dot{\epsilon}$ and $\ddot{\epsilon}$. It was even more difficult to obtain independent adjustments of the multiplying factors for $\dot{\epsilon}$ and $\ddot{\epsilon}$. Therefore, no practical results can be reported at this time in regard to high-order optimum response systems using linear switching.

Solar-Powered Thermoelectric Generator Design Considerations

N. F. SCHUH
MEMBER AIEE

R. J. TALLENT
ASSOCIATE MEMBER AIEE

Synopsis: With the coming space age, considerable attention is being focused upon the problem of supplying secondary power to the vehicles of the future. It is anticipated that future space vehicles may perform missions which will require months and perhaps years to complete. Requirements for the secondary power source serving these vehicles will be very stringent. Weight will be at a premium, therefore, the unit must have a high power-to-weight ratio. Reliability must be high but must be accomplished with little or no maintenance and without excessive redundancy. The life expectancy must be high.

Present-day systems possess a number of shortcomings which may preclude their use in extended mission vehicles. There is some doubt that these systems can be successfully modified to overcome these shortcomings. Accordingly, much effort is being expended to explore new types of power systems which will be able to meet the needs of this new age.

One system that appears to show considerable merit is a solar-powered thermoelectric generator. Solar power is fairly abundant in space, and the solar-powered system would have an advantage in that it would require no fuel or shielding. The thermoelectric generator is a static conversion device and should have a fairly long life with good reliability. This paper presents, in a brief manner, some of the principles and problems which may be expected in applying solar energy to a thermoelectric generator serving a space vehicle and also describes a small solar-powered thermoelectric generator which was constructed to study these problems.

A THERMOELECTRIC generator operates as a heat engine to statically convert heat into electricity. As such, heat must flow through the thermoelectric material. Some of this heat is converted into electricity and the remaining unavailable energy must be dissipated. A solar system will thus consist of:

1. A means for collecting and concentrating the solar energy.
2. A means for converting the radiant energy into heat.
3. A thermoelectric generator.
4. A means for dissipating the waste heat.

Basic Elements

Collector–Concentrator

The solar-energy density at various points in the solar system is shown in

Paper 59-847, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 23, 1959.

N. F. Schuh is with Westinghouse Electric Corporation, Lima, Ohio, and R. J. Tallent is with the Boeing Airplane Company, Seattle, Wash.

The authors wish to acknowledge the valuable guidance of S. J. Angelo and others, Westinghouse Research Laboratory, in the field of thermoelectricity; and the contributions made by the members of the Materials and Processes Staff and Systems Management Office, Boeing Airplane Company, in the field of solar energy.

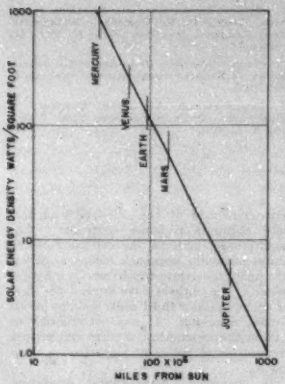


Fig. 1. Solar energy density at various points in solar system

Fig. 1. This plot shows that the power densities in the areas of space in which man may be interested will be too low to efficiently operate a heat engine so some means must be provided for increasing the energy density to workable levels.

The energy density can be increased by collection and concentration with some type of a reflective concentrator or mirror. The optical design and construction of a large-scale reflective concentrator for a power system will be similar to that of a solar furnace. However, concentration requirements will be much less. The term "concentration ratio" as used hereafter is defined as the ratio of the image-energy density with concentration to the energy density without concentration assuming a perfect concentrator with a reflectance of unity. Table I shows the input heat density of some present-day thermoelectric materials and the concentration ratio necessary to provide these densities in space. Note that only fairly modest ratios will be required, whereas solar furnaces usually require concentration ratios in the order of 10,000–30,000. A review of the published papers[1–14] covering the theory and design of solar furnaces indicates that with the low concentrations required, the optical design of a power-system concentrator should not be a major problem.

Table I. Input Heat Density of Some Present-Day Thermoelectric Materials

Thermoelectric Materials	Heat-Input Density, Watts/Sq Ft	Concentration Ratio Required to Supply Heat Input Plus Radiation Loss
Lead–telluride—P Lead–telluride—N	1,400	26
Zinc–antimonide—P Indium–antimonide—N	5,000	61
Lead–telluride—P Indium–antimonide—N	4,000	51

Solar constant = 130 watts/sq ft
Hot junction temperature = 450 C (723 K)
Cold junction temperature = 270 C (543 K)
Length of element = 0.5 inch
$\alpha = \epsilon = 1.0$
Furnace factor = 0.8
Solar-image area equal to load area

In space applications the major concentrator problem is one of a mechanical nature. Because of the large area which will be required, the collector must be in a collapsed form during the ascent through the dense air and must then be extended or erected upon reaching space altitudes. The extension system must be capable of forming a large complex surface with good accuracy. Considerable work is being done by the National Aeronautics and Space Administration and others on the problem of developing suitable extension systems.

Some shapes may be more easily formed by an extension system than others. Some reflective systems which have been considered for terrestrial applications are shown in Fig. 2 and may be worth considering from the standpoint of reducing the extension problem. Fig. 3 compares the concentration ratio of some of the more basic designs. Note that for the small concentration ratios, the simple shapes perform nearly as well as a parabola.

In a practical system, the concentrator will not deliver all of the intercepted energy to the solar image. Losses will occur due to the absorption at the reflective surface, scattering as caused by diffuse reflection, and imperfections in the basic shape. The total intercepted energy must be increased by some factor to account for these losses. In solar furnace work this factor is commonly called the "Furnace Factor" and was found to vary from about 0.3 to 0.7. Because of the low concentration ratio required in a thermoelectric system, it is expected that somewhat higher factors may be attained.

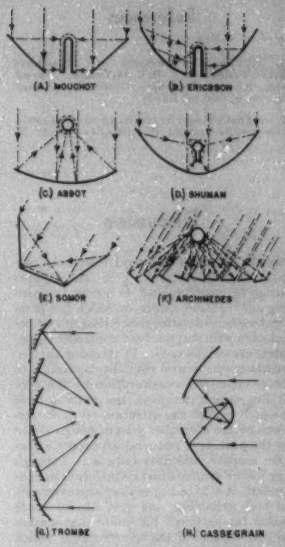


Fig. 2. Solar energy concentration systems, from references 1–3

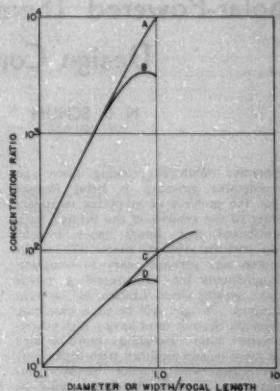


Fig. 3. Concentration ratios for various concentrator designs

A—Paraboloid or circular Archimedes array
B—Semisphere
C—Parabolic cylinder or rectangular Archimedes array
D—Semicircular cylinder

Absorber

The absorber has the function of converting the collected radiant energy into heat and conducting this heat to the generator. The conversion efficiency will depend upon the shape, surface materials, area, and temperature. The effect of the absorber shape on performance will be discussed in a companion paper.[4]

The incident energy which strikes the absorber will be either reflected or

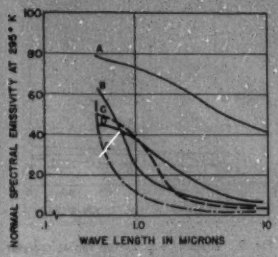


Fig. 4. Spectral emissivity of some elemental surfaces; data taken from reference 5

A—Graphite
B—Tantalum
C—Tungsten
D—Iron
E—Copper

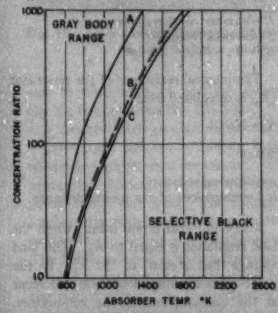


Fig. 5. Dividing lines determining application areas for absorber surface coatings

A—$\alpha = 0.8$, $\epsilon = 0.1$
B—$\alpha = 0.7$, $\epsilon = 0.5$
C—$\alpha = 0.4$, $\epsilon = 0.07$
Solar constant = 130 watts/ft²
Furnace factor = 0.8
Reference gray body = $\alpha = \epsilon = 0.95$

absorbed. The ratio between the absorbed energy and total-incident energy defines the absorptivity coefficient of the surface. The absorbed heat will thus be equal to

$$Q_a = SA_c F\alpha$$

where

α = absorptivity coefficient to solar energy
S = solar energy density, watts/sq ft (watts per square foot)
A_c = area of the concentrator, sq ft
F = furnace factor

Some of the absorbed heat will be lost by infrared radiation from the hot absorber surface. This loss can be expressed by the following equation:

$$Q_r = \epsilon\sigma A_a (T_a^4 - T_s^4)$$

where

ϵ = emissivity coefficient at the temperature of the absorber
σ = Stefan-Boltzman constant = 5.267×10^{-9} watts/sq ft/degree⁴
A_a = area of the absorber, sq ft
T_a = temperature of the absorber, degrees Kelvin (K)
T_s = effective sink temperature, K

The heat available to the generator will thus be the difference between the absorbed power and the radiated heat. The heat-balance equation of the absorber can be written as follows:

$$Q_a = Q_g + Q_r$$

or

$$\alpha S A_c F = Q_g + \epsilon\sigma A_a (T^4 - T_s^4)$$

where

Q_g = heat delivered to the thermoelectric generator

From this equation, it is evident that the heat delivered to the thermoelectric generator will be increased when either the absorptivity is increased, the emissivity decreased, or the absorber area decreased by increasing the concentration ratio. By maximizing the power delivered, the area of the concentrator can be reduced and a saving in weight realized.

If the absorber area is made small in an effort to reduce the radiation loss and the input heat density becomes much larger than the generator density, much of the heat must then be conducted over a fairly long path. The weight of the heat transfer system may offset the weight saving in the concentrator. Additional study will be required to optimize the absorber area with respect to the other elements of the solar system.

It is obvious that for a given absorber area, concentrator, and solar constant, the heat delivered will be a maximum when the solar absorptivity is equal to unity and the low-temperature emissivity is equal to zero. Unfortunately, materials with such properties are not available. However, in many materials the emissivity-absorptivity coefficient is not a constant but changes with temperature and wavelength. It may be possible to select materials which have the property of efficiently absorbing solar energy but inefficiently radiating infrared thus approaching the ideal material. Such materials are called selective blacks. Fig. 4 shows the spectral emissivity characteristics for some materials.

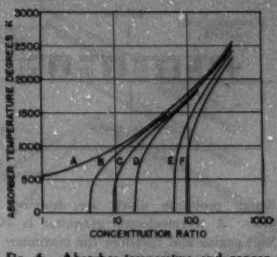


Fig. 6. Absorber temperature and concentration ratio relations at various input heat densities to thermoelements

A—0 watts/ft²
B—250 watts/ft²
C—500 watts/ft²
D—1,000 watts/ft²
E—2,000 watts/ft²
F—5,000 watts/ft²
Solar constant = 130 watts/ft²
Furnace factor = 0.7
$\alpha = 0.6$, $\epsilon = 0.1$

From the absorber heat-balance equation, it can be shown that under certain conditions, selective black materials with less than ideal characteristics will be superior to highly absorptive gray-body materials ($\alpha = \epsilon$ = constant). Fig. 5 compares three selective blacks with a highly absorptive gray body. These curves show the concentration ratios and temperatures at which a selective black is equal to the gray body. The selective black will be superior at all points falling to the right of the appropriate curve and the gray body superior for all points to the left side of the curve. In general, for a given concentration ratio, the selective blacks will be superior at the higher temperatures and the gray body at the lower temperatures.

The absorber heat-balance equation also provides a means for calculating the temperature of the absorber. The temperature of the absorber is jointly fixed by the heat loss from the absorber surface and the heat flow through the thermoelectric generator to the radiator and thence to space. The expression defining the absorber temperature in these terms is quite complex and awkward. The calculation can be simplified by assuming the power flow through the thermoelectric generator to be a constant. Based on this simplifying assumption, Fig. 6 shows the absorber temperature as a function of concentration ratio with input heat density as a parameter.

Thermoelectric Generator

A brief review of thermoelectric theory may be helpful in understanding the

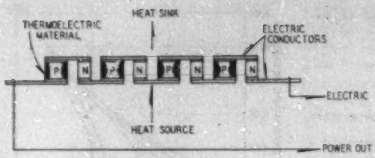


Fig. 7. Basic thermoelectric generator arrangement

design problems which are described later. A thermoelectric generator is a heat engine and therefore the maximum available fraction of the total energy is given by the Carnot cycle efficiency. This cycle efficiency, E_c is expressed as the difference in the heat-in and heat-out temperatures divided by the heat-in temperature in absolute degrees.

$$E_c = \frac{T_a - T_r}{T_a}$$

The portion of this available energy which can be converted by a thermocouple (called the thermocouple efficiency, E_{tc}) is a function of the thermoelectric material properties and the temperature. The basic material properties are the Seebeck coefficient S (in volts-per-degree-centigrade temperature difference between junctions), thermal conductivity, k, and electrical resistivity, ρ. A figure of merit Z, relating these materials properties is expressed as $Z = S^2/\rho k$.

By considering the effects of Peltier heat, Thomson heat, and Seebeck voltage it can be shown that the thermocouple efficiency is

$$E_{tc} = \frac{\sqrt{1+TZ}-1}{\sqrt{1+TZ}+T_a/T_r}$$

Now the over-all efficiency E of a thermoelectric generator can be expressed as $E = E_c E_{tc}$ and the electric power output available is $P = Q_g E_c E_{tc}$, where Q_g = the heat flow into the thermoelectric material.

It is important to note that the material properties appear only as S^2 divided by ρk in the expression for efficiency. This indicates that the material properties are not independently important but only in respect to one another.

Metals are not good materials for thermoelectric power generators because of their low thermocouple efficiency. By plotting the figure of merit with respect to the current-carrier density or resistivity it can be shown that the best thermoelectric materials apparently lie in the semiconductor or semimetal range. Indeed, most of the material development today is directed at thermoelements in this range. Unfortunately, semiconductors are not high-temperature materials, being limited to 500–700 C (degrees centigrade) as far as we know today. This means that the heat energy available in the form of burning fuels, concentrated sunlight, and nuclear reactors cannot be used at its maximum temperature and therefore maximum energy conversion cannot be obtained. It would be very desirable then to increase the efficiency by increasing the operating temperature of the thermoelement materials. A large effort is underway to convert insulator class materials into good thermoelectric materials by increasing their electrical conductivity without appreciably increasing their thermal conductivity or decreasing their Seebeck voltage. Mixed valence materials such as lithium nickel oxide appear very promising in the 700–1,500 C range.

The major electrical characteristics of a thermoelectric generator can be defined by the power output, the open-circuit voltage, and the internal resistance. The power output at matched load has previously been given approximately as the product of the heat flow into the material and the over-all efficiency. The open circuit voltage (V_o) is calculated from

$$V_o = NS\Delta T$$

where

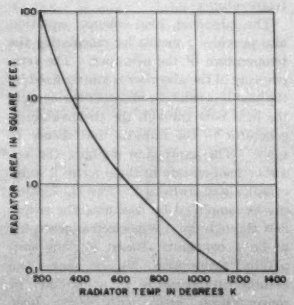


Fig. 8. Radiator area required to dissipate 1 kw of heat

$\epsilon = 1.0$
Sink temperature = 0 K

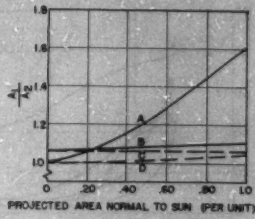


Fig. 9. Effect of exposure to sun on radiator area requirements

A—$\epsilon = \alpha = 0.95$, $T_r = 500$ K
B—$\epsilon = 0.9$, $\alpha = 0.1$, $T_r = 500$ K
C—$\epsilon = 0.9$, $\alpha = 0.1$, $T_r = 1{,}000$ K
D—$\epsilon = \alpha = 0.95$, $T_r = 1{,}000$ K
A_1 = area required with various exposures to sun
A_4 = gray-body area required with no exposure to sun

N = number of thermoelements
S = Seebeck voltage
$\Delta T = T_a - T_r$

The internal resistance of the generator (R_g) is calculated from

$$R_g = N\rho l/A$$

where

ρ = electrical resistivity
l/A = dimensions of the thermoelements

Maximum load power occurs when the load resistance is equal to R_g. The load voltage then is equal to 1/2 V_o. In the practical case voltage drops will also occur at the contacts and in the conductors. These drops are generally assumed to be no greater than 5% of the total voltage.

The design of a thermoelectric generator will vary greatly with energy source, application, environment, and material state-of-the-art. Whatever the details of appearance, however, a heat path must be provided from the heat source, through the thermoelements, and into the heat sink as shown in Fig. 7. Generally speaking, the elements are thermally in parallel and electrically in series which indicates one of the problem areas in any practical design. Thermal contact areas must be good heat conductors and at the same time must be good insulating electrical surfaces.

Another problem area results from the relatively high operating temperatures encountered. High-temperature thermoelements must be connected together with high-temperature conductors, held together with high-temperature structural materials, and protected by high-temperature finishes. These requirements reduce the number of available materials which

Fig. 10. Solar-powered thermoelectric generator test model, vacuum enclosure not shown

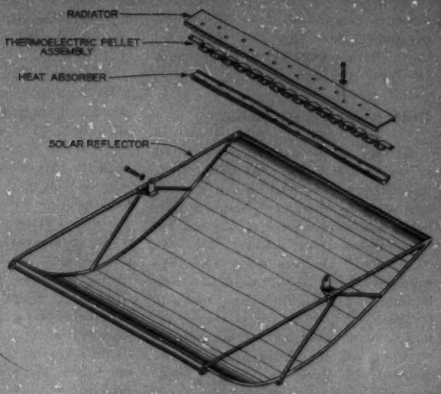


Fig. 11. Exploded view of principal elements of solar-powered thermoelectric generator test model

can be used even at nominal thermoelectric temperatures like 400–500 C.

A thermoelectric generator is essentially a low-voltage high-current device with a relatively high internal impedance. This means that the output voltage will change considerably with load. The change in load voltage with load current will not be exactly linear because of the Peltier effect. The Peltier effect is in such a direction as to tend to cool the hot junction and heat the cold junction. As the current is reduced by increasing the load resistance, the hot and cold junction temperatures tend to increase. These increases tend to be approximately the same so that ΔT is almost constant. Deviations from linear voltage regulation are due to changes in ΔT and the resultant changes in material characteristics with temperature change.

Because of the inherently slow response time of such a thermally operated device, good voltage regulation under load changes would be difficult to obtain by trying to control the heat flow. In most applications it is anticipated that some form of power-conversion equipment will be used either to raise the direct voltage level or to convert the direct current to alternating current. Voltage regulation can be very nicely accomplished in this conversion equipment.

Radiator

The function of the radiator is to dissipate the waste heat from the generator unit and to control the temperature of the cold junction. At the present, the radiator appears to be one of the larger problem areas in a solar-powered thermoelectric system.

In space, the only mode for heat removal is by radiation. Fig. 8 shows the familiar relation between the radiator area and temperature required to dissipate 1 kw of heat. To attain a high Carnot efficiency, the radiator temperature should be as low as practicable, however, low temperatures can only be attained with large radiator areas. With a large radiator and concentrated input, a major portion of the waste heat will be dissipated at areas somewhat removed from the source, and some means must be provided for conveying the heat to these areas.

If a solid-material heat-conduction system is used, the flow of heat by conduction to the dissipative areas will cause a temperature drop. If the drops are large, extra area must be added to compensate for the reduced performance. Since this area will be added at an even greater distance from the source, the dissipative capabilities will be quite low and a point of diminishing returns is rapidly reached.

The flow of heat in a radiator is a complex problem of two-dimensional heat flow. This problem is currently being studied with the aid of a computer. At the time of writing, detailed results have not been obtained. Approximate results indicate that if the power inputs are high and radiation temperatures low, fairly large thermal-conductor cross sections will be required with attendant larger weights.

In a space solar-powered system, the efficiency of the conversion device may not be as significant as in the more familiar terrestrial applications. Solar energy is abundant and free. The main consequence of reduced efficiency will be to increase the size of the concentrator and the conversion device. Since the goal in space applications is to obtain a large power-to-weight ratio, it may be desirable to sacrifice efficiency by using higher radiator temperatures, thereby reducing the radiator weight. Optimization studies will be required to determine the best radiator temperature.

Another factor which will influence the radiator area is the orientation of the emissive surfaces. In some solar systems, the radiator will be exposed to the sun. If the radiator area is to be kept to a minimum, the heat input from the sun must be kept small. This can be done by keeping the area facing the sun small and by using suitable selective black material for the radiator surface. Fig. 9 shows the effect of selective blacks on the total area required with various percentages of the area normal to the sun. Note that if good selective blacks are available or if the temperature is high the solar energy does not significantly increase the radiator area.

In other solar systems, the radiator will be shaded from the sun but exposed to the earth (for earth-orbiting vehicles) during certain portions of the orbit. The earth may be considered as a black body radiating at the average temperature of the earth and reflecting a certain percentage of the solar radiation. Calculations indicate that the effect on radiator require-

ments is insignificant at the higher temperatures. From this work, it appears that radiator orientation need not significantly affect radiator design.

Description of Test Model

A review of the considerations covered in the foregoing indicated that certain problem areas will exist which are not amenable to analytical solution. To better define and evaluate these problems, a solar-powered thermoelectric model was constructed. Temperatures and outputs were established by material and size limitations. No attempt was made to optimize any part of the system or to make the model flyable. The primary objective was to construct a workable model which could be easily instrumented.

The power output was arbitrarily set at 2.5 watts. It was estimated that a minimum thermocouple efficiency of 7.7% could be achieved with today's material and that a Carnot cycle efficiency of 25% was reasonable. This gives an over-all efficiency of 1.92%. Based on this figure a heat input to the thermoelements of 130 watts is required.

The complete model is shown in Fig. 10 and as an exploded view in Fig. 11. It consists of a concentrator, a sidereal drive, an evacuate enclosure, a water-cooled sink and the thermoelectric assembly.

Concentrator

Preliminary investigation on the problem of model concentrator construction and orientation indicated that a cylindrical type mirror would have some advantage over the other types. Such a mirror could be constructed by metal-forming techniques rather than grinding and would require accurate orientation in only one plane. It was thus decided to construct a nonflyable parabolic mirror of this type. The dimensions were arbitrarily set at a width of 20 inches and an aperture of 50 inches by the requirement that the mirror or concentrator be large enough to collect a demonstrational quantity of energy yet small enough to be transportable. The proposed mirror intercepts 6.93 sq ft of solar energy. With a solar constant of 75 watts/sq ft, this is equal to 520 watts.

The focal length was arbitrarily set at 12.5 inches to make the system compact. With this focal length and neglecting aberrations, the concentrator should focus the solar image in the form of a rectangle with a width of 0.116 inch and a length of 20 inches. The aberrations associated with a concentrator of such high aperture diameter to focal length increases the width beyond the theoretical 0.116 inch.

According to catalog information highly reflective aluminum sheets were available which had a solar radiation reflectivity of about 0.83. This material appeared to be equal to, or better than, other finishes with the exception of silver and would not be subject to discoloration from atmospheric corrosion. This material was thus selected as the reflective surface.

The mirror was constructed by mounting the highly polished aluminum sheet in a welded tubular frame. The parabolic surface was generated by preforming the aluminum sheet and inserting it into two slots that were accurately milled in the frame. The mirror and frame assembly was then mounted upon a sidereal drive.

Thermoelectric Calculations

A hot junction temperature of 450 C was selected based on considerations of material availability and long life. With a Carnot efficiency of 25% the cold junction would then be 270 C. Thermoelectric materials such as p-type zinc antimonide and n-type indium antimonide were available and have useful values of efficiency over this temperature range.

The decisions of electric power output, thermoelectric material, and operating temperatures having been made, the component parts could now be designed. The amount of thermoelectric material required is calculated by considering the thermal conductivity and electrical resistivity of the materials and the heat flux desired. The ratio of area to length of the p-type material is calculated using the optimum relationship of the material characteristics.

$$Q_p = \left[1+\frac{\sqrt{\hat{\rho}_n \hat{K}_n}}{\sqrt{\hat{\rho}_p \hat{K}_p}}\right] K_p \frac{A_p}{l_p}(T_a - T_r)$$

$$130 = \left[1+\frac{\sqrt{(5.7)(10^{-4})(9.2)(10^{-3})}}{\sqrt{(2.0)(10^{-3})(2.5)(10^{-3})}}\right]\times (2.5)(10^{-3})\frac{A_p}{l_p}(450-270)$$

$$\frac{A_p}{l_p} = 14.9 \text{ cm (centimeters) or } 5.9 \text{ inches}$$

In a similar manner, the n-type material is found to require an A/l of 4.1 cm or 1.6 inches. Heat flow being dependent on the ratio of area to length, theoretically the dimensions could be reduced to near zero and the same heat flow obtained as long as the ratio is maintained. In practice the minimum length is limited by mechanical considerations such as thermal stress and heat loss around the elements.

An arbitrary length of 1/2 inch was chosen for both types of material to simplify construction. This is not necessarily the optimum length but rather a convenient one. Setting the length makes it possible to calculate the total areas as:

$A_p = 3.0 \text{ inches}^2$

$A_n = 0.8 \text{ inch}^2$

The selection of diameters for the individual thermoelements is determined by mechanical considerations and the output voltage required. It is generally desirable to break up the total area of thermoelectric material required into as many separate junctions as possible to increase the voltage output.

In this case the p-type material was readily available in 0.5-inch-diameter pellets. The number of p-type pellets then is:

$$N_p = \frac{3.0}{\pi 0.25^2} = 15.5$$

or

N_p = approximately 16

To obtain the same number of matching n-type pellets:

$$r_n^2 \pi = \frac{0.8}{16}$$

and

$r_n = 0.13$ inch

Therefore the n-type pellets are approximately 0.25 inch in diameter.

The open-circuit voltage then is $V_o = (16)(158+160)(10^{-4})(180) = 0.92$ volts; and the internal resistance is $R_g = 0.07$ ohms. The power output at matched load then is $P = 3$ watts. The calculated 3-watt output indicates that the over-all efficiency is slightly higher than that assumed originally. In practice, however, contact drops, conductor resistance, and mechanical deviations will tend to reduce the output below this calculated value.

Nickel-plated silver was chosen for conductor material because of its high electrical and thermal conductivity and its corrosion resistance. The radiator was made of aluminum because of its light weight and high thermal conductivity.

An aluminum oxide coating was added to provide electrical insulation between radiator and thermoelement conductors. The radiating surface was coated with a black paint to provide maximum radiation of the waste heat.

The absorber was made from a low-carbon steel. This material has an ac-

ceptable thermal conductivity but more important it has a matching coefficient of linear expansion. The expansion of the steel absorber at its higher temperature matches very closely the expansion of the aluminum radiator at its lower temperature. This arrangement minimizes the lateral stresses on the thermoelements. The round side of the semicylindrical absorber was coated with tantalum so that it absorbs a majority of the concentrated sunlight and radiates very little heat. The flat side was coated with aluminum oxide to provide electrical insulation and thermal conduction to the thermoelement conductors. Stainless steel hardware was selected to hold the structure together. Stainless steel is corrosion free and has a low thermal conductivity thus minimizing the loss of heat through the hardware. Spring washers were used to take up dimensional changes caused by thermal expansion.

Absorber

It was shown that 130 watts at a temperature of 450 C must be supplied to the thermoelectric elements to produce an electric power output of 2.5 watts. In addition, mounting and hardware conduction losses will exist. The total net power which must be supplied to the generator is calculated to be 140 watts.

As described in a foregoing section, the heat available to the generator is determined by the absorber heat-balance equation.

$$Q_a = Q_g + Q_r = \alpha A_c S F = Q_g + A_c \epsilon \sigma (T_a^4 - T_s^4)$$

In this case, the term Q_g is equal to the heat input into the thermoelements plus mounting and hardware losses which is equal to 140 watts. The intercepted energy $A_c S$ has already been established by the dimensions of the concentrator and was found to be 520 watts with a solar constant of 75 watts/sq ft. An absorber area must now be found which will balance the equation with the absorber at the desired temperature of 450 C.

The absorptivity–emissivity coefficients are a function of the surface coating. Preliminary work indicated that a selective black material may be superior to a highly absorptive gray-body material in this application. A brief search of the literature revealed tantalum to be one of the better blacks and it was selected as the surface material. The solar absorptivity was calculated to be about 0.5 and the emissivity to be about 0.11.

A furnace factor of 0.83 appeared reasonable in view of the high reflectivity of the concentrator surface. The assembly will be mounted in an evacuated enclosure and some of the concentrated energy will be reflected and absorbed in passing through the glass walls. A transmission factor should thus be added to the left term in the heat-balance equation. Calculations and tables indicate that the transmission factor should be about 0.9.

The required absorber area of 31.7 square inches is now calculated by substituting these values into the heat-balance equation and letting $T_s = 323$ K (50 C). Since the radiation losses from the back side of the absorber will be reduced to a small value with insulation, all radiation losses must occur on the front surface. With the length of the absorber fixed at 20 inches by the concentration dimensions, the required width is now found to be 1.59 inches. This width is larger than the theoretical 0.116-inch image that will be cast by the concentrator. However, certain inaccuracies in construction are expected and it is felt that a sizable target will be desirable to effectively collect the concentrated energy. The dimension was thus left as calculated.

From preliminary work, it appears that a plane surface will not efficiently collect energy from the outer surfaces of the concentrator. This is due to the large incident angle of these rays. A cylinder will present a surface area that will be approximately normal to all the rays and should be superior. The absorber was thus constructed as a semicylinder with a diameter of 1 inch.

Evacuated Enclosure

A vacuum enclosure was considered necessary to eliminate convection heat loss and to protect the materials from possible oxidation damage. The envelope consists of a 4-inch Pyrex tube with mechanically sealed end caps. Electrical connections to the thermoelectric generator, voltage probes, and temperature sensing thermocouples were brought out through hermetically sealed connectors mounted in the end caps. The end caps also contain connections for the water supply to the heat sink and the vacuum system. The enclosure is attached to the concentrator frame with an adjustable type of clamp to permit positioning at the point of maximum heat input.

Sink

The Pyrex walls of the enclosure are opaque to the wavelengths emitted by the radiator. The enclosure would thus absorb all of the waste heat and reach an objectionably high temperature. It was considered simpler to absorb the waste energy in an internally mounted water-cooled sink rather than to attempt to cool the glass tube. The sink consists of a semicircular section formed from thin-gage copper which was blackened on the concave side and aluminized on the convex side. Copper tubes for the coolant were soldered to the copper sheet at approximately 2-inch intervals.

Radiator

The heat which must be dissipated by the radiator is equal to the heat conducted into the thermoelectric elements and generator hardware less the electrical output. In this case, the heat input to the radiator is 127.5 watts from the thermoelectric generator plus 10 watts hardware conduction loss for a total of 137.5 watts.

The familiar Stefan-Boltzman radiation law provides the basis for sizing the radiator.

$$Q_w = A_r \epsilon \sigma (T_r^4 - T_s^4)$$

where

Q_w = waste heat to be dissipated = 137.5 watts
A_r = area of radiator in sq ft
ϵ = effective emissivity = $(1/\epsilon_r + 1/\epsilon_s - 1)^{-1}$
ϵ_r = emissivity coefficient of radiator surface
ϵ_s = emissivity coefficient of sink surface
T_r = radiator temperature = 543 K
T_s = sink temperature = 288 K

Since the radiator is shaded from the sun by the sink, minimum area will be obtained with highly emissive surfaces. Emissivity tables show that surfaces with emissivities in the order of 0.9 can be obtained with lampblack carbon.

The required area of the radiator was calculated to be 56.7 square inches. The radiator was made as a plate 2.8 inches wide and 20 inches long. For structural reasons, the thickness of the radiator could not be made less than 1/8 inch. With power inputs of approximately 70 watts/ft and material thickness of 1/8 inch, the conductor temperature drop should be small, and no allowance is considered necessary to account for varying radiator temperatures.

Test Program

The test model will be subjected to a series of tests designed to determine the performance of the system and to investigate the optical, thermal, and electrical problems. Theoretical considerations will be proved or disproved or otherwise evaluated by comparing calculated results with experimental data. The experience gained in designing, constructing, and testing this small model will be

of considerable benefit in future solar-thermoelectric work.

Conclusions

This paper discusses some design considerations and problems that will be encountered in designing a solar-powered space thermoelectric generator. Results of some studies are presented and a test model of a solar-powered thermoelectric generator is described.

It appears that a solar-powered thermoelectric generator is feasible and may have considerable potential as a secondary power source for future space vehicles. There are many problems which must be explored but these problems appear to be solvable with further engineering effort.

References

1. Power from Solar Energy, J. I. Yellott. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 79, no. 6, Aug. 1957, pp. 1349–59.

2. High Temperature Furnaces, Felix Trombe. *Proceedings*, Phoenix Symposium on Applied Solar Energy, Phoenix, Ariz., 1955, pp. 63–72.

3. Recent Progress in Solar Furnaces for High Temperature Research and Development Work, W. M. Conn. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 257, no. 1, Jan. 1954, pp. 1–11.

4. Optimum Reflector–Absorber Geometry for a Solar Generator, R. W. Stineman. *AIEE Transactions*, see pp. 332–37 of this issue.

5. American Institute of Physics Handbook (book), D. E. Gray. McGraw-Hill Book Company, Inc., New York, N. Y., 1957.

6. Analysis of Large Aperture Parabolic Mirrors for Solar Furnaces, J. Farber, B. I. Davis. *Journal*, Optical Society of America, New York, N. Y., vol. 47, no. 3, Mar. 1957, pp. 216–20.

7. A Segmented-Mirror Solar Furnace for High Intensity Thermal Radiation Studies, R. Gordon. *Review of Scientific Instruments*, New York, N. Y., vol. 25, no. 5, May 1954, pp. 459–63.

8. Fuel for Solar Furnaces, A. R. Kasmandel. *The Journal of Solar Energy Science and Engineering*, Phoenix, Ariz., vol. 1, no. 2–3, Apr.–July 1957, pp. 44–47.

9. Designing Solar Furnaces for Specific Performance, Raymond Bliss. *Ibid.*, pp. 55–62.

10. Solar Energy Furnaces, Guy Beauvineis, Stanford Research Institute, Menlo Park, Calif., 1955.

11. Semi Conductor Thermoelements and Thermoelectric Cooling (book), A. F. Ioffe. Infosearch, Ltd., London, England, 1957.

12. Climatological Data. Weather Bureau, U. S. Department of Commerce, Washington, D. C., vol. 8, no. 1–12, Jan. to Dec. 1957.

13. Process Heat Transfer (book), Donald Q. Kern. McGraw-Hill Book Company, Inc., 1950.

14. Heat Transmission (book), William H. McAdams. McGraw-Hill Book Company, Inc., 1942.

Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components

W. B. SISCO
NONMEMBER AIEE

J. M. FISKIN
NONMEMBER AIEE

VEHICLES traveling at hypersonic velocities in a gaseous atmosphere will be surrounded by a shock-induced high-temperature sheath. As a result of the high temperature, the gaseous sheath has an electron density greater than that of the ambient atmosphere. Since the craft will be immersed in a high-temperature envelope, both it and its components will be subjected to unusual temperatures. Fig. 1 indicates the way in which the air temperature and density varies from the shock front to vehicle surface for a given velocity and altitude.

Currently designed aircraft have temperatures up to 1,000 K (degrees Kelvin) over a majority of the surface. This raises the temperature of interior components to 1,000 K unless a special design is used. As a consequence, emphasis is being placed upon the design of high-temperature electrical and electronic components or on means to protect these components from the hot environment.

In addition to the high-temperature effects, ionization itself directly affects systems on the vehicle, on other vehicles, and on the surface of the earth.

On the aircraft the ionized sheath surrounding the vehicle may cause the following results:

1. Malfunction or inefficient operation of electronic equipment due to the deviation of the radiation resistance from free-space values.

2. Increase in the open-circuit noise in receiver systems.

3. Change in the voltage breakdown properties from ambient properties.

4. Modification of propagation characteristics of the media.

At locations distant from the vehicle two changes will occur. These are:

1. A variation in the radar cross section of the craft.

2. Enhanced response to passive detection devices due to radiation from the hot gases.

Temperature Effects on Equipment

Initially, the design of high-temperature electrical and electronic systems will be based on the prevention of the hot environment from reaching the components. One approach to this is to place the equipment in cooled compartments with only the components able to function at high temperatures external to the compartments.

The design of the compartment so that it will perform its protective function becomes important. One such arrangement is shown in Fig. 2. Essentially, the result of such a configuration is to make it necessary to remove only the heat rejected by the operating equipment and the heat transferred from 373 K [212 F (degrees Fahrenheit)] to 310 K (100 F) through the insulation. This can be done by refrigeration or by blowing cool gases over the equipment.

Many electrical components such as isolation transformers, wiring, coaxial cables, waveguides, etc., will probably be located at points in the vehicle that are inaccessible to practical cooling. As a result, they will have to function at temperatures up to 1,000 K. The design of these components will have to rely on materials having the required electrical properties such as dielectric constants, dielectric strengths, conductivities, etc, at these high temperatures. Present indications are that this has been accomplished in many cases. The liberal use of ceramics as dielectric and structural material has increased the high-temperature capability of many passive components. The white noise in the system will increase with the temperature, how-

Paper 59-885, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing April 22, 1959.

W. B. Sisco and J. M. Fiskin are with Douglas Aircraft Company, Inc., Long Beach, Calif.

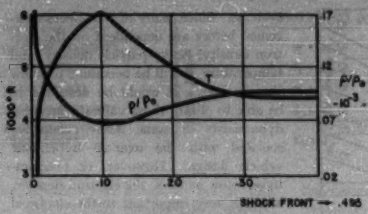


Fig. 1. Temperature and density ratio at various distance from surface

Altitude = 260,000 feet
ρ_0 = density at standard pressure and temperature
U_1 = 25,000 fps
X = 9 feet aft of L.E.

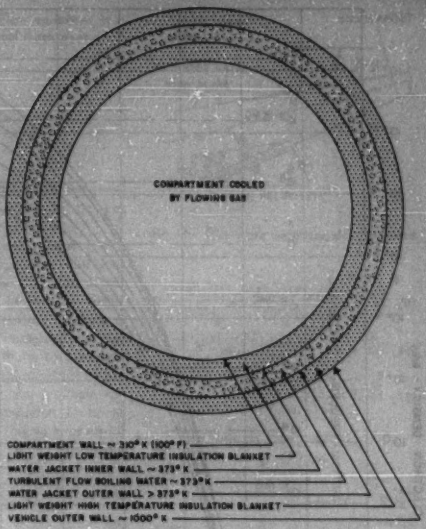

COMPARTMENT COOLED BY FLOWING GAS
COMPARTMENT WALL ~ 310° K (100° F)
LIGHT WEIGHT LOW TEMPERATURE INSULATION BLANKET
WATER JACKET INNER WALL ~ 373° K
TURBULENT FLOW BOILING WATER ~ 373° K
WATER JACKET OUTER WALL > 373° K
LIGHT WEIGHT HIGH TEMPERATURE INSULATION BLANKET
VEHICLE OUTER WALL ~ 1000° K

Fig 2 (right). Example of a protected compartment

ever, and the high-temperature system's signal-to-noise ratio will decrease.

Ionization Effects on Equipment

Radiation Impedance

Because the dielectric constant of the medium surrounding the vehicle is not equal to that of free space, the radiation resistance of antennas attached to the surface of the vehicle can differ significantly from their free-space values. These variations will not necessarily remain constant with time. Unless this radiation-resistance variation is compensated, it could result in the malfunction or inefficient operation of electronic equipment. These impedance changes can be obtained approximately by classical EM (electromagnetic) theory provided the geometry of the sheath is known.

Noise Impedance

All of the vehicular antennas are surrounded by a conductive medium at relatively high temperatures. The real impedance, as differentiated from the radiation impedance, appearing across the antenna terminals will depend on the geometry of the antenna feed points, the shape of the shock-wave sheath and the conductivity and dielectric constant of the plasma. The resistive part of this impedance, which can be considered in parallel with the modified radiation resistance previously discussed, will generate an open-circuit noise voltage over a wide-frequency spectrum. If there are no relatively strong EM fields present within the medium, the noise power will be proportional to the electron temperature, i.e., the temperature of the shocked air. Strong EM fields, for instance, those produced by high-power radar systems, yield effective electron temperatures significantly above the thermal values resulting in even larger noise-power outputs.[1] Generally, the noise in the receiver systems will be increased as a result of the hot gases surrounding the vehicle.

Voltage Breakdown

High-temperature air is basically different from air at standard pressure and density. The extent of the difference depends on the temperature which determines the quantitative amounts of dissociation, ionization, and formation of new products taking place. It is to be expected then that the voltage breakdown properties would differ from ambient air. Theoretically, the breakdown properties could be obtained since the composition of the air is known. Practically, the calculation of ionization and excitation rates is difficult and experimental methods are usually employed.[2] The problem is thus one of maintaining the air at the desired temperature and density for the experiment.

Radar Cross Section

Air and ground detection and tracking systems will also be affected by the ionization present in the sheath surrounding the vehicle. Fig. 3 is a plot of the electron density and plasma-resonance frequency as a function of temperature for various ratios of shocked-air density to standard temperature and pressure-air density. Fig. 4 is a combination of Figs. 1 and 3 showing the way that the plasma-resonance frequency, which is proportional to the square root of the electron density, varies across the sheath.

The shock-ionized layer may be considered as an extension of the vehicle for incoming EM radiation at frequencies in the neighborhood or lower than the critical frequency of the plasma. Additionally, if recombination times are long, an ionized wake may also be present. This, in effect, changes the shape and the radar cross section of the vehicle for some frequencies. Again, this variation can be approximated if the shape of the sheath and wake are known.

Radiation from the ionized sheath may be roughly classified as classical and quantal. Accelerations and deaccelerations of the electrons within the gas will yield radiation similar to that produced in antennas. Rotation, vibration, electronic excitation, and recombination of electrons with positive ions will produce quantal radiation in the microwave, millimeter, infrared, visible and ultraviolet ranges. Thus, the tracking of the vehicle by passive detection systems will be enhanced.

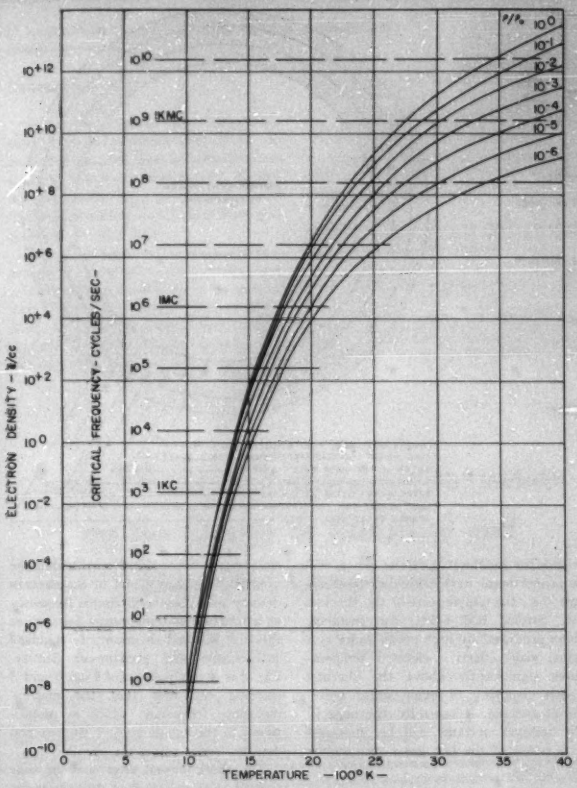


Fig. 3 (above). Critical frequencies for high-temperature air

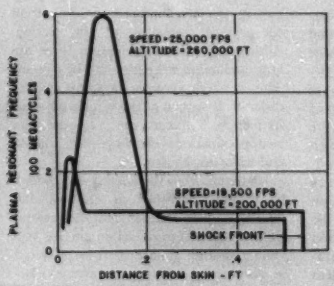


Fig. 4 (left). Plasma resonant frequency at various distances from surface, 9 feet aft

Basic Design Parameters

Treating the ionized sheath as an ionosphere surrounding the vehicle, the propagation and reflection characteristics of the medium can be determined by employing classical EM theory.[3-4]

Some of the terms used previously are not generally thought of as design terms to be applied to electrical and electronic systems. Since many of the effects mentioned before are dependent on the electron density, heavy-particle density, and temperature, it will be necessary to know these values. It would be desirable to be able to obtain those values from aerodynamicists or some other group associated with the over-all hypersonic vehicle design. However, at temperatures below 5,000 K the electron density, which is very important to the electrical and electronic systems, is generally a negligible factor in their design considerations. Therefore, the electrical engineer is faced with learning a new field in order to obtain design requirements. It is desirable to review briefly the principal theory by which the high temperature and electron density are determined.

Composition and Energy of Gas Behind a Shock Wave

For most aerodynamic shapes the equilibrium composition of the gas and the temperature from the shock front up to the boundary layer (a narrow layer adjacent to the skin of the vehicle) can be obtained. Under some conditions, the boundary-layer composition can also be calculated. The necessary fundamental concepts and equations will be briefly outlined in the following. A more detailed account can be found in reference 5.

From an aerothermodynamic standpoint the problem is essentially one of conserving mass, momentum, and energy flow across the shock front. Expressed mathematically, conservation of mass flow is:

$$\rho_0 U_0 = \rho_1 U_1 \tag{1}$$

conservation of momentum flow is:

$$P_1 - P_0 = \rho_1 U_0 (U_0 - U_1) \tag{2}$$

and conservation of energy flow is:

$$E_1 + P_1 \tau_1 + \frac{1}{2} U_1^2 = E_0 + P_0 \tau_0 + \frac{1}{2} U_0^2 \tag{3}$$

and

$$h_1 - h_0 = \frac{1}{2}(P_1 - P_0)(\tau_1 + \tau_0)$$

where

ρ = density (mass/unit volume) and $1/\rho$ (specific volume) = τ
U = streamwise flow velocity normal to shock front
P = pressure
E = internal energy/unit mass (function of T, temperature)

Zero subscripts indicate the unshocked or ambient gas; a one signifies the shocked gas.

The shock front, which separates the ambient air from the shocked air, can

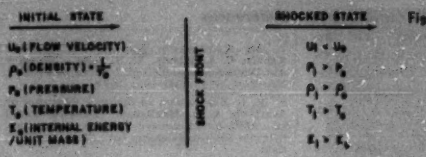


Fig. 5. Flow equations in a normal shock wave

be considered as a transformer in that it changes the temperature, density, pressure, internal energy, and streamwise velocity to different values. These transformations are conveniently shown in Fig. 5. It is noted that all the values, with the exception of the streamwise velocity, increase as the gas flows through the shock front. This, in effect, means that energy in the streamwise flow is converted into internal energy and compression of the gas. An idealized picture of how this streamwise flow energy is converted into the various modes of internal energy is shown in Fig. 6. In Fig. 7 the geometry of a shock wave surrounding a flat surface is shown.

Equations 1, 2, and 3, in conjunction with the perfect gas law,

$$P_1 v_1 = \Sigma n_j k T_1 \quad (4)$$

where

n_j = number in volume τ of species j
k = Boltzmann's constant

then can be solved for a vehicle traveling at a specified altitude and velocity provided the composition and the number of energy modes present remain the same on both sides of the shock front. In the earth's atmosphere, for air, this would mean that dissociation and ionization do not occur and vibrational and electronic states are not excited. For even relatively low hypersonic velocities, however, the temperature T_1 of the shocked air is sufficient to dissociate, ionize, and excite the particles in the gas. To satisfy the conservation equations under these conditions, the composition of the gas behind the shock must be known.

The composition (number density) of a gas at a specified temperature and density or pressure can be determined employing the principles of physical chemistry and statistical quantum mechanics. The first step in the numerical calculation of the concentration is to postulate the chemical reactions taking place within the hot shocked air. As a simple example, if air were considered to be at 3,000 K or below, the following group of equilibrium relations would generally be valid.

$$N_2 \rightleftarrows N+N \quad (5)$$

$$O_2 \rightleftarrows O+O \quad (6)$$

$$NO \rightleftarrows N+O \quad (7)$$

$$NO \rightleftarrows NO^+ + e^- \quad (8)$$

Equations 5 and 6 are usually sufficient to determine the aerothermodynamic properties of the gas. However, since the electron density is the primary requirement, equations 7 and 8 must also be included. The inclusion or exclusion of an equation depends on whether or not it will significantly affect the aerothermodynamic properties or the number density of the particular species required.

Associated with each chemical equation is an equilibrium constant, K, which can be related to the concentrations and partition functions of each particular species by the law of mass action. Thus,

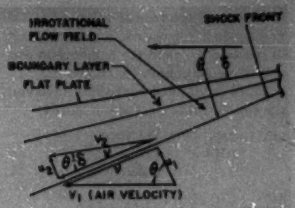


Fig. 7. Flat plate aerodynamic conditions

$$K_5 = \frac{[N][N]}{[N_2]} = \frac{(PF)_N (PF)_N}{(PF)_{N_2}} \left(\frac{1}{\rho_{dr} L}\right) e^{-\frac{Q_5}{kT}} \quad (9)$$

$$K_6 = \frac{[O][O]}{[O_2]} = \frac{(PF)_O (PF)_O}{(PF)_{O_2}} \left(\frac{1}{\rho_{dr} L}\right) e^{-\frac{Q_6}{kT}} \quad (10)$$

$$K_7 = \frac{[N][O]}{[NO]} = \frac{(PF)_N (PF)_O}{(PF)_{NO}} \left(\frac{1}{\rho_{dr} L}\right) e^{-\frac{Q_7}{kT}} \quad (11)$$

$$K_8 = \frac{[NO^+][e]}{[NO]} = \frac{(PF)_{NO} + (PF)_{e^-}}{(PF)_{NO}} \times \left(\frac{1}{\rho_{dr} L}\right) e^{-\frac{Q_8}{kT}} \quad (12)$$

where

K = equilibrium constant
ρ_{dr} = ratio of actual density to standard density
L = number of air atoms at standard density
k = Boltzmann's constant
Q = energy of dissociation, formation, and ionization
[ ] = concentration (ratio of number of atoms of a particular species, such as N, to total number of atoms present)

The total partition function (PF) is a function of T and can be written very generally as:

$$(PF) = (PF)_{\text{translational}} \times (PF)_{\text{rotational}} \times (PF)_{\text{vibrational}} \times (PF)_{\text{electronic}} \quad (13)$$

where the partition functions for the various energy modes can be obtained from statistical quantum mechanics[8] and

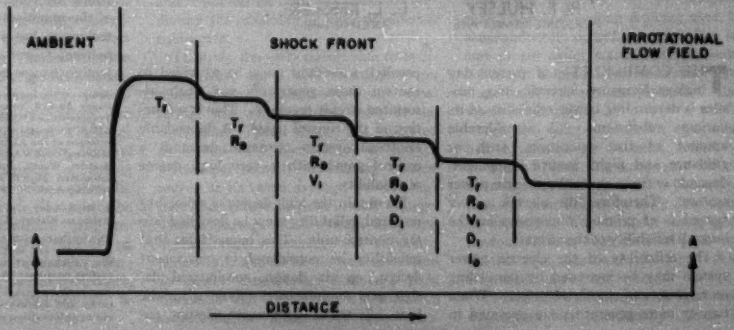


Fig. 6. The shock front cross section

T_t = translation
R_o = rotation
V_i = vibration
D_i = dissociation
I_o = ionization

experimental spectroscopic data.[7-11] The (PF), when calculated, and the energy of reaction, Q, yield the equilibrium constants in equations 9, 10, 11, and 12.

Three additional equations involving the concentrations are also required.

$$2[N_2]+2[O_2]+2[NO]+2[NO^+]+[N]+[O]+[e^-]=1.000 \quad \text{mass conservation} \quad (14)$$

$$\frac{2[N_2]+[NO]+[NO^+]+[N]}{2[O_2]+[NO]+[NO^+]+[O]}=4.0$$

material balance conservation–ratio of nitrogen to oxygen atoms (15)

$$[NO^+]=[e^-] \quad \text{charge conservation} \quad (16)$$

The seven equations 9, 10, 11, 12, 14, 15, and 16 form a set containing seven unknowns, the concentrations, []. The simultaneous solution of this set yields the desired concentrations and hence the composition of the gas.

The total energy of the shocked gas including the amount necessary to dissociate, form, and ionize can be obtained once the composition and (PF)s are known. Thus,

$$E=E_{part}+(E_{diss}-E_{form})+E_{ion} \quad (17)$$

Here E_{part} is the total internal energy of the particles and includes translational, vibrational, rotational, and electronic modes and is given by:

$$E_{part}=\sum_{fi} n_{fi}\frac{\partial}{\partial T}\ln (PF)_f \quad (18)$$

where f identifies the species of the particle (i.e., N, O_2 . . .) and i the electronic energy level of the species. The dissociation energy is simply the number of particles that dissociate times the dissociation energy Q. Identical procedures are used for the ionization and formation energies.

In obtaining the composition the shocked-air temperature and density were assumed to be known. Actually, they can only be obtained by satisfying the flow conservation relations, (equations 2, 3, and 4). The flow-conservation equations can only be satisfied if the composition is known. To resolve this problem, some type of an iterative procedure is usually adopted. Probably the easiest method in most cases is to compute the composition over a wide range of shocked-air temperatures and densities and then use an iterative procedure to satisfy the conservation equations for a range of ambient velocities and altitudes (densities). The composition tables of Gilmore[10, 12] have been recently extended to include electron densities down to 1,000 K.

Conclusions

Some effects of temperature on equipment and the equipment's environment have been presented. Special emphasis has been placed on a new field of knowledge required by the electrical engineer in the design of systems for hypersonic aircraft.

References

1. Air Composition Below 3000 Degrees Kelvin, C. A. Roberts, W. B. Sisco. "Research and Development Handbook," *Space/Aeronautics*, New York, N. Y., 1959–60, p. D-19.

2. R. F. Breakdown in Air as a Function of Altitude, A. D. MacDonald. *Technical Report no. MPL-2*, Microwave Physics Laboratory, Sylvania Electric Products, Inc., Mountain View, Calif.

3. Electromagnetic Waves and Radiating Systems (book), E. C. Jordan. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1950, pp. 660–62.

4. Electromagnetic Wave Propagation in Shock-Ionized Media, W. B. Sisco, J. M. Fiskin. *Engineering Paper no. 633*, Douglas Aircraft Company, Inc., Santa Monica, Calif., Apr. 1958.

5. Theory of Equilibrium Electron and Particle Densities Behind Normal and Oblique Shock Waves in Air, C. A. Roberts, W. B. Sisco, J. M. Fiskin. *Transactions*, Professional Group on Antennas and Propagation, Institute of Radio Engineers, New York, N. Y. (To be published.)

6. Statistical Mechanics (book), G. S. Rushbrook. Oxford University Press, London, England, 1949, chaps. IV, XI, XII.

7. Spectra of Diatomic Molecules (book), Gerhard Herzberg. D. Van Nostrand Company, Inc., Princeton, N. J., second edition, 1950.

8. Atomic Spectra and Atomic Structure (book), Gerhard Herzberg. Dover Publications, Inc., New York, N. Y., second edition, 1944.

9. Atomic Energy Levels. *Circular no. 467*, National Bureau of Standards, Washington, D. C., vols. 1 and 2, 1949, 1952.

10. Equilibrium Composition and Thermodynamic Properties of Air to 24,000° K (USAF Project Rand), F. R. Gilmore. *Report RM-1543*, The Rand Corporation, Santa Monica, Calif., Aug. 1955.

11. Photodetachment Cross Section and the Electron Affinity of Atomic Oxygen, Lewis M. Branscomb, et al. *Physical Review*, New York, N. Y., vol. 111, no. 2, July 15, 1958, pp. 504–13.

12. Effects of Relatively Strong Fields on the Propagation of EM Waves Through a Hypersonically Produced Plasma, W. B. Sisco, J. M. Fiskin. *Transactions*, Professional Group on Antennas and Propagation, Institute of Radio Engineers, July 1959, pp. 240–44.

Functional Cycling to Assure Reliability of Aircraft Control Equipment

R. E. HULSEY
ASSOCIATE MEMBER AIEE

L. L. KESSLER
AFFILIATE AIEE

THE COMPLEXITY of present-day high-performance aircraft and missiles is demanding better reliability of its various subsystems. A considerable amount of the equipment, such as guidance and flight control equipment, depends entirely upon the electric power system. Therefore, the electric power system is of primary importance in the over-all reliability of the aircraft.

The reliability of the electric power system may be increased by paralleling more than one source of power. When two or more generators are operated in parallel, a method must be provided to control these generators and their associated circuit breakers. This is a function of the control panel. A dependable electrical system therefore demands a control panel with a very high degree of reliability.

To obtain the high degree of reliability required, reliability must be designed into the control unit. This necessitates that reliability be considered in component design, circuit design, mechanical design, and manufacture. Even when the best efforts of component design engineers, manufacturers, and quality assurance personnel are used, studies show that early component failures cannot be eliminated. An effective method has been developed for eliminating these premature failures from static control and protective panels by functional cycling at the end of the manufacturing process. This new technique has increased the mean-time between failure of static control and protective devices by 5,000 per cent.

Paper 59-912, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing April 28, 1959.

R. E. Hulsey and L. L. Kessler are with Westinghouse Electric Corporation, Lima, Ohio.

The authors wish to acknowledge the work performed by Mr. R. G. Lee and Mr. D. A. Vogel in the development and design of the functional cycling equipment described herein.

The authors wish to thank the aircraft manufacturers and operators for their co-operation in providing information on equipment in the field.

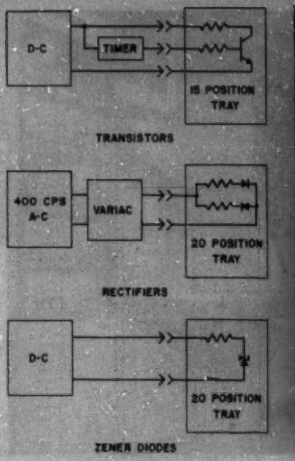


Fig. 1. Block diagram of functional cycling equipment for semiconductors

Fig. 2. Functional cycling equipment for semiconductors as described in Fig. 1

Typical Control Panel

The basic functions of a control and protective panel are those of opening and closing the field of the generator and supervising the opening and closing of the generator circuit breaker. In a parallel generator system, supervising the opening and closing of the bus tie circuit breaker is another basic function. In order to perform these control functions in a logical manner, a control and protection panel must sense voltages and currents throughout the electrical system. These various voltages and currents are measured, compared with standards, time delayed if necessary, mixed in logic circuits, amplified, and finally used to perform the necessary control functions. The panel protective functions may include such circuits as overvoltage protection, undervoltage protection, overexcitation protection, underexcitation protection, underspeed protection, underfrequency protection, stability protection, negative sequence protection, and differential current protection. The panel control functions may include lockout circuits to prevent cycling of circuit breakers and a circuit for automatic paralleling of generators. To meet the size, weight, environmental and reliability requirements of present-day high-performance aircraft, circuits utilizing static components, such as transistors, control rectifiers, and diodes are used in place of electromechanical devices.

Verifying the Need for Functional Cycling

It is well known that electronic components exhibit the characteristic of a high initial failure rate. In connection with the development of new products utilizing semiconductors, extensive studies were conducted on semiconductors to determine their initial failure rate characteristics. As a part of these studies, extensive life tests of semiconductors were conducted. In order to predict the reliability of production control equipment with some degree of accuracy, life tests on representative quantities of semiconductors were necessary. Reliability studies on pilot production components and equipment were inconclusive and could not be used as a basis for predicting the reliability of later production equipment.

To study the early failure rate of semiconductors, tests were conducted on representative quantities of each type, rating, and manufacturer used in the equipment. Each diode was operated as a half-wave rectifier on a 400-cps (cycles per second) alternating current source at its rated average forward current and maximum peak inverse voltage. The average forward voltage drop and average reverse leakage was measured periodically and all diodes were monitored to detect catastrophic failures. Each transistor was operated as a switch with rated collector voltage when not conducting, and rated collector current while conducting. During the tests each transistor was operated conducting half, and not conducting half, of each hour. The leakage, direct current gain, and saturation resistance were measured periodically and all transistors were monitored to detect catastrophic failures. Each semiconductor was operated for a minimum of 200 hours.

The results of the studies showed the characteristic of a high initial failure rate. The majority of early catastrophic failures occurred during the first 100 hours of operation. These results were presented to semiconductor manufacturers and arrangements were made to age all semiconductors prior to delivery. It is impractical to age all semiconductors electrically, therefore, extreme temperature aging and cycling is being used by semiconductor manufacturers. To assure proper aging, samples of each semiconductor shipment are electrically aged (functional cycled) as described elsewhere in this paper.

The first pilot production equipment had a mean-time between failure of 80 hours. Since this equipment did not have aged semiconductors, a low mean-time between failure in early life would be expected. To eliminate these early random failures, each piece of pilot production equipment was functional cycled for a minimum of 200 hours.

Production equipment with aged semiconductors was functionally cycled for

Fig. 3 (left). Punch card controlled component tester. Zener diodes are being tested

Fig. 5 (right). Functional cycling equipment for control panels as described in Fig. 4

comparison with pilot production equipment. This comparison showed a considerable reduction in failures but random failures still occurred. Further investigation showed that some early failures are inevitable as a result of manufacturing processing even when extreme care is exercised. The failures are of the nature that cannot be detected visually or by short time functional tests. Therefore, the only satisfactory means for eliminating these random failures is to functional cycle each unit for a minimum of 100 hours. Data from the functional cycling test are continually surveyed to establish improved procedures.

Functional Cycling Equipment

Many factors must be considered in the design of functional cycling equipment. The design of the equipment must be flexible, in that it can be readily adapted to new types of control equipment. It is desirable that it be built in module form so that it may be easily maintained. The equipment should be as automatic as is economically possible, providing unattended operation and including self-checking features. Due to the great number of operations that this equipment must perform, it must be designed with the greatest reliability attainable. Because of the differences between semiconductor component and control equipment cycling, two basic types of functional cycling equipment were required.

The functional cycling equipment for semiconductor components is shown in Figs. 1, 2, and 3. It was found convenient to divide the semiconductor components into three categories for the cycling equipment. These categories are transistors, rectifiers, and zener diodes. All components are mounted in plug-in trays for cycling. These trays can be removed from the cycling racks and inserted into automatic test equipment which tests all components on the tray automatically.

Transistor cycling is accomplished by applying a constant direct-current voltage supply through a resistor to the collector continuously. The transistor is conducting half of each hour and is controlled by a timer operated relay through a base-drive limiting resistor. Rectifier cycling is accomplished by applying a 400-cycle alternating-current

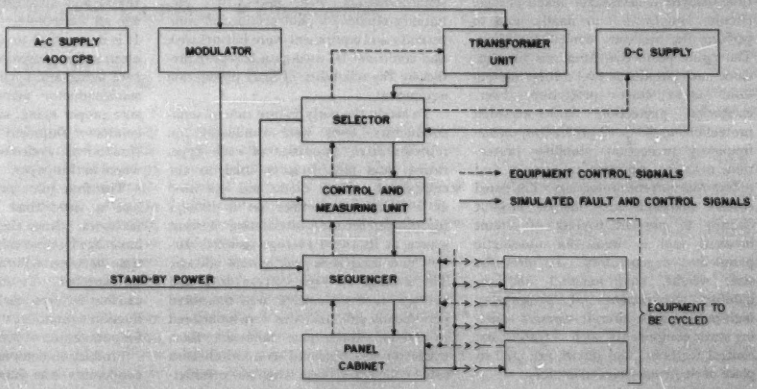


Fig. 4. Block diagram of functional cycling equipment for control panels

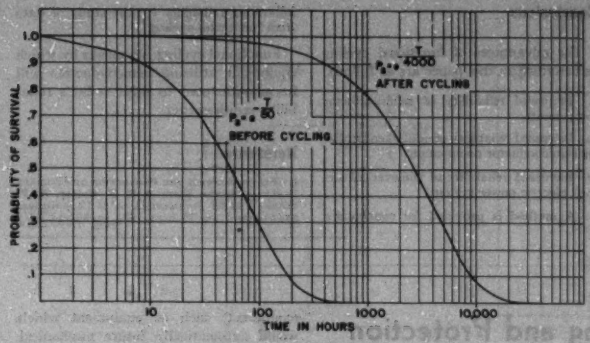


Fig. 6. This curve demonstrates the improvement in control panel reliability as a result of functional cycling

supply through a load resistor to each diode. Each plug-in tray has a variac for adjusting peak inverse voltage and is connected so half of the diodes conduct on each half-cycle. Zener diode cycling is accomplished by applying a constant direct-current voltage supply through a load resistor to each zener diode, the level depending upon the zener voltage. These trays are very flexible since all current limiting resistors are mounted on the plug-in trays.

Each tray of semiconductors is checked by automatic testing equipment at the end of each 24 hours of cycling. The testing procedure for the automatic equipment is controlled by punched cards. This equipment scans each position within the plug-in tray. If a bad component is found, the scanner stops and indicates the position of the bad component. This equipment checks a plug-in tray of 20 components in less than 1 minute.

The functional cycling equipment for the complete control panel is more complex since many control and protective functions must be cycled. A block diagram of the functional cycling test set is shown in Fig. 4. For purposes of discussion, the functional cycling equipment will be divided into four categories. These categories are power supplies, control and measuring devices, sequencer, and control panel mounting.

At the present time there are three types of power supplies required for cyling static control panels. A direct-current constant-voltage power supply, a 3-phase 400-cps constant frequency and voltage supply, and a 3-phase modulated supply obtained by modulating the 3-phase constant-frequency power supply. This modulating device can be controlled to provide amplitude modulation between zero and 40%. The various 3-phase alternating-current voltages required are obtained by using two transformers and an auto transformer for each phase. The secondaries of these transformers are tapped in such a manner that any voltage between zero and 199.5 volts line-to-neutral can be selected in half-volt steps.

The control and measuring device is the heart of the functional cycling equipment. This device controls the operation of the other units in the test set. A typical control panel requires approximately 120 individual steps for cycling. This control and measuring device will perform the 120 individual steps in less than $2^1/_2$ minutes. This control and measuring device is composed of a master control unit and a selector programmed for each particular control panel. The master control contains the basic bridges, the time delays and the means for stepping the selector. The selector is programmed to choose the basic bridge, complete this bridge, select the proper terminals of the control panel, and select the proper power supply. Different programmed selectors may be easily connected to the master control unit.

The sequencer is used to connect the control and measuring device to a particular panel. The sequencer records each complete cycle on a counter, transfers the panel to stand-by power until the next cycle and steps the control and measuring device to the next panel. The panels which fail on any individual step during the cycle are completely disconnected from the two power sources and the failure is indicated by means of a lamp on the sequencer. The sequencer is timed in such a manner that each panel is functionally cycled once each half hour regardless of the number of positions used. If desired, a panel may be selected and cycled manually. Associated with each sequencer is a cabinet containing a quick-eject rack with the proper connector for the panel to be cycled. One complete cycling unit is shown in Fig. 5.

Results of Functional Cycling

Functional cycling is provided today's aircraft and missiles with the most advanced and reliable control and protective panels ever used. Users of a typical control and protective panel have accumulated 16,632 panel-hours to date. Failures have occurred, but the mean-time between failures is over 4,000 hours compared to 80 hours prior to functional cycling (see Fig. 6). The functional cycling equipment is presently providing reliability data at a rate of 650,000 semiconductor hours per month. These data are being accumulated and compiled on a computer. Continued study of the results of functional cycling is providing information for improving component and control panel design and manufacture. These improvements will continue to increase the mean-time between failures of control and protective panels. There is a 10,000-hour life test program in progress on a particular 2-generator parallel system to further study the increased reliability resulting from functional cycling. Table I is a tabulation of data from a

Table I. Mean-Time Between Failures for Different Cycling Periods

Period (Hours)	Cumulative Hours Per Panel	No. of Panels Involved	Cumulative Panel Hours	Period MTBF (Hours)	Remarks
0 to 25	25	83	2,075	520	First 25-hour cycling period (panels had cycled components)
26 to 50	50	83	4,150	415	Second 25-hour cycling period (panels had cycled components)
51 to 75	75	83	6,225	415	Third 25-hour cycling period (panels had cycled components)
76 to 100	100	83	8,300	2,075	Fourth 25-hour cycling period (panels had cycled components)

representative sample of one type of control panel. These data on 83 panels were accumulated during the 100-hour cycling period. The table shows a significant increase in the mean-time between failure during the fourth 25-hour period as compared to any previous 25-hour period. A 100-hour cycling period is adequate for this panel. Another panel, depending on its complexity and on the reliability of the components, may require a cycling period of a longer or a shorter duration.

Conclusions

The advantages of functional cycling as described in this paper are as follows:

1. Increased reliability of semiconductor components.
2. Increased reliability of equipment utilizing semiconductor components.
3. Increased reliability of the over-all aircraft electric power system.
4. A method is provided for continually improving reliability by the study of data from functional cycling.

Functional cycling is a means by which equipment utilizing semiconductors will be manufactured to meet today's reliability requirements.

Reference

1. Static Control and Protection of Aircraft Electric Power Systems. R. E. Hulsey, N. F. Schuh. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, Sept. 1957, pp. 206–12.

High-Speed Restarting and Protection of Large Synchronous Motors

C. L. PHILLIPS
MEMBER AIEE

M. H. YUEN
ASSOCIATE MEMBER AIEE

A WESTERN refinery was adding a new catalytic reformer unit to existing facilities. The chemical process involved, which had a platinum catalyst worth $250,000 in suspension, was dependent upon mechanical power to a compressor. This was a vital drive where loss of mechanical power for a short time meant loss of the catalyst as well as loss of production.

The use of a steam-driven compressor backed up by a steam or electric motor driven stand-by unit is a well-tried solution to such a situation. However, adequate capacity was not available from the existing facilities, and the addition of the necessary boiler capacity meant a capital expenditure of about $300,000.

Hence, the question was asked: Could power continuity be achieved electrically? Two separate reclosing sources of power were to be supplied by the utility and the user assumed the risk of an extended simultaneous loss of power on both utility feeders. After a cost study of the various combinations of electrical drives, three synchronous motors were selected. All three were needed for peak production, but only one was necessary to preserve the catalyst; this represented a double stand-by at the drives. The question now became: What would be the additional cost to achieve short-time continuity of the distribution system between the two power sources and the three drives? Cost estimates indicated that this could be done for between 10% and 20% of the steam stand-by, and the decision was made to go ahead.

The decision to use synchronous motor drivers with a goal of short-time continuity brought with it a series of considerations involving motor protection. Considerable concern has been centered about possible physical damage to large motors where a substantial residual voltage and high-speed restarting and/or a reclosing source are coincident. (See the references.)

Hence, two programs exist: First, high-speed restarting to achieve process continuity and, second, adequate motor protection. This paper demonstrates, with an oscillographic record of successful operation, an electrical solution of these two coincident problems.

System Layout and Equipment

Fig. 1 is the refinery electrical 1-line diagram. Two separate source feeders with reclosers are both energized but only one source energizes the main refinery bus. If the normal utility source fails entirely, a 1.3-sec (second) throwover to the other source is actuated, re-energizing the main refinery bus. Note that both feeders to the catalytic reformer are normally closed, and that, hence, a utility power discontinuity appears as a momentary complete loss of power to the catalytic reformer. This refinery substation is modern drawout switchgear and is rated AM-13.8-250 megavolt-amperes.

Fig. 2 shows the catalytic reformer 1-line diagram. Here the intent was to provide functional duplication of all equipment, such as equipment which would automatically insure mechanical power to the compressors even though a failure of any piece of equipment or cable were to occur. Second-order failures were considered but not provided for in all cases.

The new 4.16-kv drawout switchgear is rated AM-4.16-150 megavolt-amperes. It is fed by two 1,500-kva transformers rated 12-kv delta to 4.16-kv Y, in one line-up. A 4.16-kv bus is connected overhead to a Limitamp starter line-up of conventional design. The synchronous motors all are rated 600-horsepower, 1.0 power factor, 400-rpm, 4,000-volt, 69.3-ampere full load, 50-degree-centigrade temperature rise, force-ventilated, with separate motor-generator set excitation and 8,500 lb-ft^2 (pound-feet^2) inertia. The compressor is a positive-displacement horizontal opposed-compression frame with 1,000 lb-ft^2 inertia. The 480-volt auxiliaries are fed by a conventional double-ended load center with each end rated 750-kva, 12-kv delta to 480-volt delta. Conventional magnetic drawout air circuit breakers were used at the 480-volt level with a 1-sec time delay throwover between the main secondary breakers and the bus tie breaker. This throwover combined with a lack of under voltage protection for the individual exciter motor-generator sets assures excitation power if there is voltage from either 480-volt switchgear bus. Conventional motor control centers were used. Without exception, all equipment was of outdoor weatherproof construction.

Figs. 3 and 4 respectively illustrate the

Paper 59-832, recommended by the AIEE Chemical Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing May 4, 1959.

C. L. Phillips is with the General Electric Company, San Francisco, Calif., and M. H. Yuen is with the Bechtel Corporation, San Francisco, Calif.

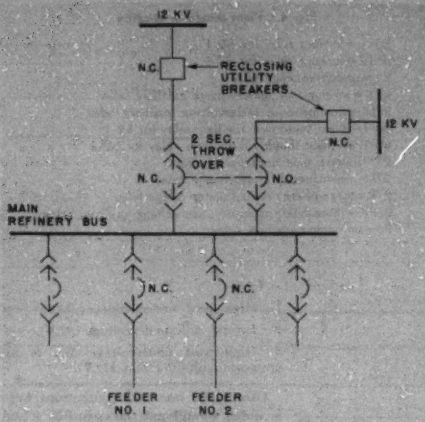


Fig. 1 (above). Over-all refinery 1-line diagram

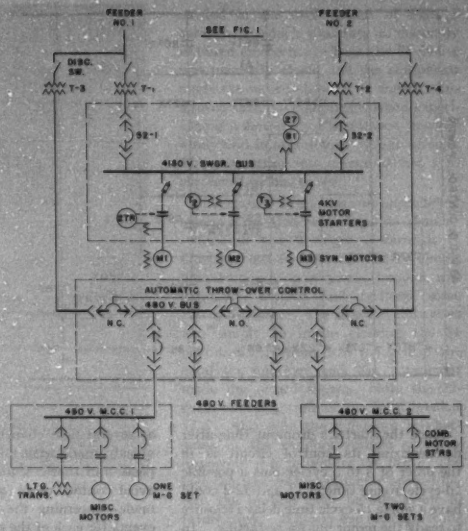


Fig. 2 (right). Catalytic reformer 1-line diagram

trip-and-close circuits of *52-1* and *52-2*, and the simplified plant shutdown circuit. The philosophy of this relaying is based on power continuity being maintained to these synchronous motors at any cost short of destroying the motors.

An important risk to the running motor from a high-speed reclosure of a source is twisting a shaft or distorting a winding. The best assurance against such a failure, at this writing, is to remove, somehow, the motors from the bus before a reclosure can re-energize the motors. This was done with a relay tripping *52-1* and *52-2* after the initial opening of the source breaker. The relay chosen here was a General Electric Company type *CFF13A* high-speed underfrequency (*81*) relay. It was originally set at 58 cps (cycles per second).

Once *52-1* and *52-2* are open, conditions on both sides of these circuit breakers must be returned to normal as rapidly as possible in order to initiate restarting. The reclosure and/or throw-over upstream will restore system voltage to normal anywhere within 20 cycles to 5 sec after the initiating disturbance.

After an open circuit between the source and the motors is established, the motors act as synchronous generators; therefore the same auxiliary relay *86-1*, which is initiated by the *81* that trips the *52-1* and *52-2*, also de-energizes the motor fields. This action is quite essential. With field applied, the motor voltage will vary almost directly as the motor speed, which will show little variation over the short time involved. Without field excitation, the motor voltage decay is primarily a function of the motor time constant, which will usually result in acceptable voltages within a time measured in cycles. By measurement from the oscillographs of Figs. 5 through 11, the time constant, represented by the time required for the armature voltage to decay from very nearly full voltage to 36.8% or 1/e pu (per unit) when the field is removed, is in this case 0.153 sec. In order to check this action, a 3-phase residual relay on the first motor to restart acts as a lockout of the starter until this motor's residual voltage decays to 25% of normal. The other two motors are restarted upon a signal from timing relays set much longer than the time required for the residual voltage to reach 25% of normal.

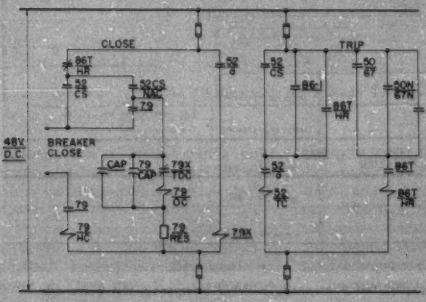


Fig. 3. Simplified breaker trip-and-close circuits of 52-1 and 52-2

52 = air circuit breaker
52/a = air circuit-breaker auxiliary contact
52/cs = air circuit-breaker control switch
52/TC = air circuit-breaker trip coil
50/67 = phase directional overcurrent relay, instantaneous and inverse time
50N/67N = phase directional overcurrent relay, instantaneous and inverse time
51 = phase overcurrent relay inverse time
86T/HR = lockout relay, hand reset
86-1 = plant shutdown auxiliary relay
79 = single-shot reclosing relay
79X = reclosing auxiliary relay

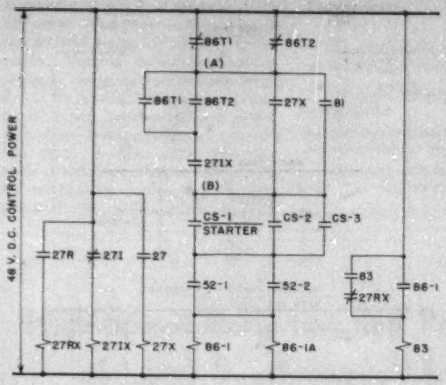


Fig. 4. Plant shutdown circuitry

86T1 = lockout relay for 52-1
86T2 = lockout relay for 52-2
81 = underfrequency relay
27X = inverse time undervoltage auxiliary relay
271X = instantaneous undervoltage auxiliary relay
86-1 = plant shutdown auxiliary relay for 52-1
86-1A = plant shutdown auxiliary relay for 52-2
27 = inverse time undervoltage relay
271 = instantaneous undervoltage relay
27R = three-phase residual undervoltage relay
27RX = three-phase residual undervoltage auxiliary relay
83 = unloading auxiliary relay

Since the starter's drop-out time after de-energizing its control circuit is in the order of 3 to 5 cycles plus a possible 10-cycle arcing time, *52-1* and *52-2* could have a single 15-cycle time delay reclosure to ready the motor bus for restarting. In designing the relaying it appeared at first as though the tripping and reclosing of the 4.16-kv breakers could be omitted in favor of simply opening the starter contactor. However, the uncertainties of contactor opening time plus a desire for positive operation resulted in the decision to use the deliberate and predictable action of the breakers. The problem can then be demonstrated with:

1. A system layout with adequate standby capability such as that illustrated in Figs. 1 and 2.
2. Breaker trip and close circuits such as that in Fig. 3, illustrating necessary manual controls, conventional overcurrent and short-circuit protection, and special tripping and closing functions.
3. A shutdown circuit such as illustrated in Fig. 4, capable of sensing those conditions which would necessitate a deliberate yet high-speed shutdown and restart.
4. A modified starter control circuit to insure necessary interlocking, early field removal, lockout against excessive residual voltage, provisions for a loaded stop and an unloaded start, and necessary manual controls.

In connection with the designing of the special shutdown circuitry, some uncertainty existed as to whether all the functions and circuits that the actual installation might require could be anticipated. Consequently, if the need for a function was in doubt, it was included. Ample auxiliary relays and contacts were included in anticipation of field modifications. This approach turned out happily as several field changes were easily and quickly made with minimum loss in expense and time. Care was necessary to avoid contact races, and decisions were made concerning the sequence of relay events for each of the multiplicity of conditions that might occur. Those who may toy with the idea of trying this should realize that it takes a great deal of thought and care; simplicity should not be anticipated on the first attempt.

The principal shutdown functions discussed in this paper are found between points *A* and *B* in Fig. 4, and are here limited to:

1. High-speed underfrequency (*81*).
2. Inverse time undervoltage (*27X*).
3. High-speed undervoltage due to an upstream fault (*86T* plus *271X*).

Other, less important, functions were included, though not shown in Fig. 4, and those attempting this work may find other functions special to their conditions. Note that in the case of a shutdown because of a fault upstream from the 4.16-kv bus, the 4.16-kv circuit breaker nearest the fault is locked out but the alternate breaker is permitted to reclose to effect rapid restart.

Regarding the Limitamp starters, care is taken to insure high-speed field removal and to insure that unloading valves have time to operate before a restart. All three starters make an independent try at restarting in time sequence to relieve the strain on a possibly faltering source.

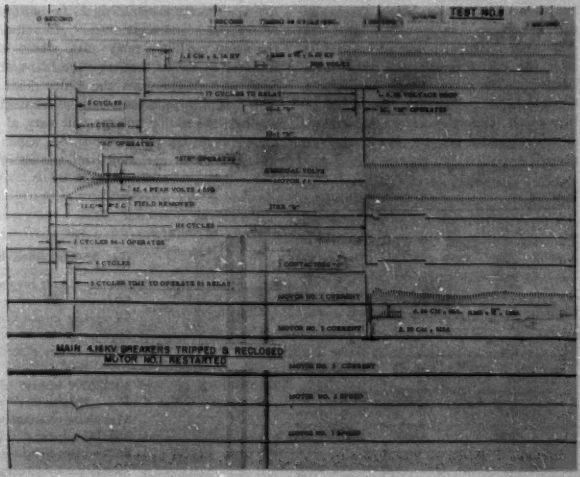


Fig. 5. Oscillographic test 8, manually initiated emergency, first 3 seconds

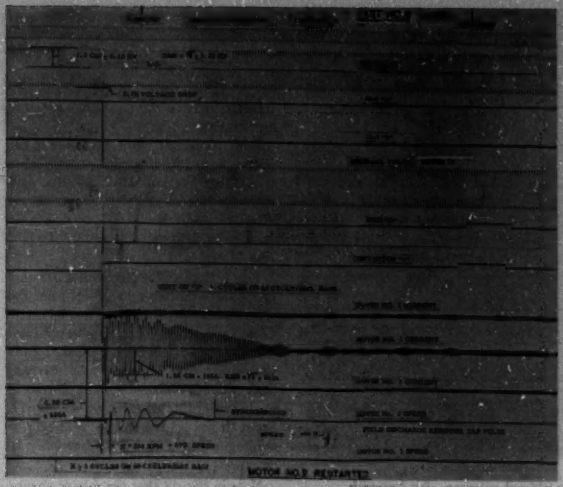


Fig. 6. Oscillographic test 8, motor 2 restarted

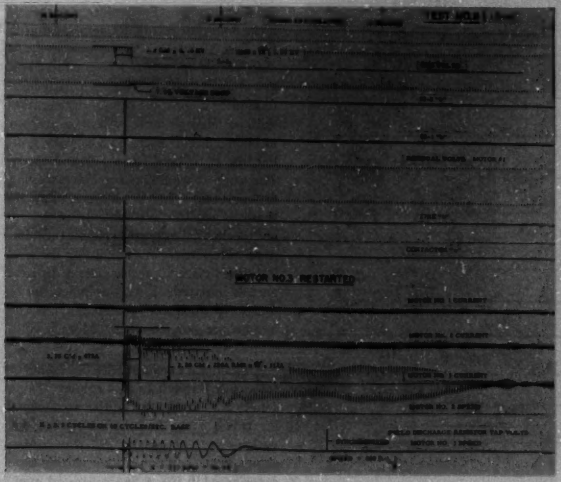


Fig. 7. Oscillographic test 8, motor 3 restarted

Another source on another system might permit simultaneous restarting with residual voltage permissive relays on all starters.

The requirements of the process are pertinent to the question of sequential starting. The subject equipment restores all three motors to synchronism within 15 sec from the moment of system disturbance. Further, this equipment could accomplish complete sequential starting within 10 sec by simple timer adjustment without stress and without compromise. This was well within the process requirement.

The equipment performance is demonstrated in the oscillograms of Figs. 5 through 7, which illustrate a manually initiated emergency. Here the *81* device was manually closed and the equipment made the record shown. It was demonstrated that the equipment, when experiencing a frequency disturbance, reacted safely and in accordance with its design. The last motor was energized and synchronized within 15 sec.

It remained to simulate an unpredicted emergency; this was done as demonstrated in test 9. Figs. 8 through 11 illustrate test 9, wherein a 12-kv circuit breaker acting as a source breaker was manually tripped and reclosed some 150 cycles later. The delay of 150 cycles was simply to insure clearing the *81* operation. Nothing was touched at the catalytic reformer and the equipment made the record shown. Note the 22-cycle reclosing time of the local breaker *52-2*. Also note the difference in character in the residual voltage and how important it is to remove the field. The fact that the motor contactor opened before the breaker may be significant in future work. If power had returned while *52-2* was open, as is possible, the *27R* would have postponed the first motor restart. The motor-starting traces on the single long oscillograph (Figs. 8 through 11) are conventional, and the voltage drops are all less than 10%. Note the over-all time of less than 15 sec and the space in the time sequence which would permit an over-all time of less than 10 sec. Motor speeds were arrived at by recording a voltage drop on a portion of the motor discharge resistor. This would show the slip frequency in the rotating field while the motor was accelerating. One slip cycle measured in X cycles on a 60-cps base gives the per-cent speed at the time of re-energizing the motor by the equation: per-cent speed $= 100\ (1-1/X)$.

Figs. 8 through 11 show that it took circuit breaker *52-2* a total of 40 cycles, on a 60-cps basis, to separate the motors from the failing source. This amount of time could be divided into 30 cycles for the *81* (set to operate at 58 cps) to operate, 2 cycles to operate the *86-1* device, and 8 cycles more to open the breaker. During this time the motors were acting as synchronous generators. Had the initial disturbance resulted in a 30-cycle reclosure, it seems clear that these motors in this system would have been exposed to the risk of shaft damage or winding distortion. The test is a good illustration of the conditions creating the possible hazards of inadequate motor protection. It is these hazards which indicated the need for an oscillographic record in the work. One

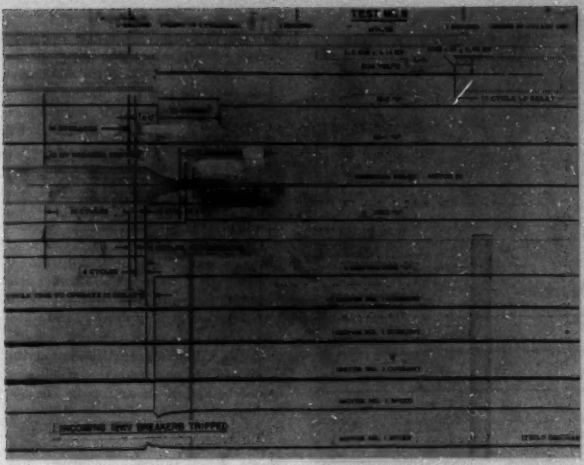


Fig. 8. Oscillographic test 9, automatically initiated emergency, first 3 seconds

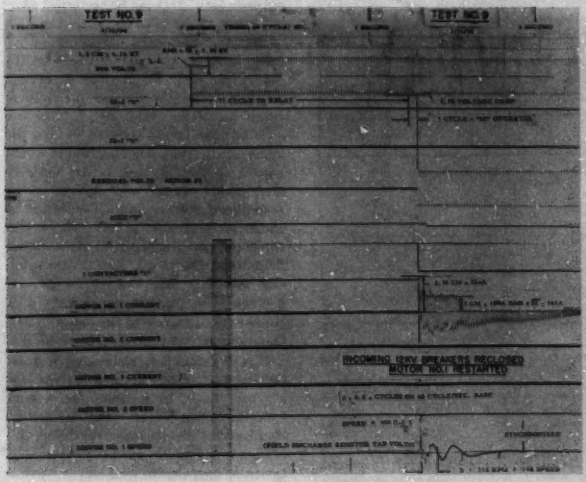


Fig. 9. Oscillographic test 9, motors 1 restarted

of the objectives of the work is an evaluation of the hazards and a reasonable assurance that they have been overcome. The authors believe that this objective was obtained in this case, as will be illustrated later in the paper.

Test 9 was initiated with no electric load in the catalytic unit other than the three synchronous motors, their exciter motor generator sets, and their cooling fans. The mechanical load on the motors was approximately 10%, consisting largely of mechanical and electrical losses in the motors and compressor. It was felt that these conditions would result in the minimum rate of deceleration that could be reasonably expected of the system. If the motors could be protected during the conditions of minimum rate of deceleration, it was believed that the system and relay characteristics would surely protect the motors during higher rates of deceleration.

For those particularly interested in the operation of the *81* device and the electrical degrees' separation of the synchronous motors and source immediately after the source breakers opened, Fig. 12 is offered. It is an enlargement of a portion of Fig. 8.

Fig. 13 depicts the electrical separation of all three motors as a group from the source after the disturbance was initiated. Common polar co-ordinates were used and the data were carefully measured from the oscillograph of test 9. Of perhaps more value are the same data plotted to log-log co-ordinates as shown in Fig. 14. From these curves 4 cps/sec is about the minimum rate of change of frequency to be expected from this system.

Note in Fig. 14 that the rate of change of frequency is shown to be constant following the first 20 cycles after initiating the disturbance. Particularly mysterious is the nonlinear rate of deceleration during the first 20 cycles. Since this was not anticipated, it received a good deal of attention. Past practice has usually been to assume a constant rate of deceleration until conditions arise where connected mechanical and electric loads vary nonlinearly at reduced frequency and speed. The authors would welcome a clarifying explanation. The significance may be that it is possible for some types of electric and mechanical systems to reduce the minimum rate of deceleration during the critical time before 20 cycles after a disturbance is initiated; if so, the difficulties of designing adequate protection for the motors are compounded. More oscillographic data are needed.

Assuming a constant rate of change of frequency throughout at 4 cps/sec, Table I indicates theoretical anticipated clearing times of the *52-2* circuit breaker for various settings of the *81* relay. Also shown is the relation of the clearing time experienced in Fig. 8 and suggested minimum source reclosing times. It is interesting to note that Fig. 15 shows a profitable gain in the *81* relay operating time up to about 10 cps/sec; this is a gain of 5 cycles. Hence, optimistic estimates of the rate of change of frequency of a particular system produce only minor reductions in total clearing times and minimum reclosing times. It can be stated, though, that a rate of change of frequency above the minimum will widen the margin in

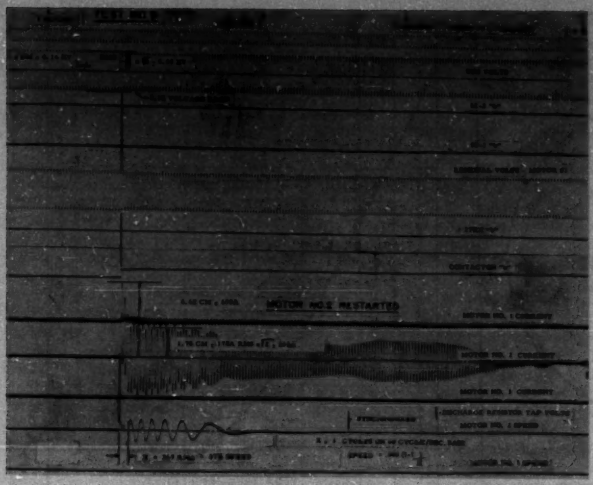


Fig. 10. Oscillographic test 9, motor 2 restarted

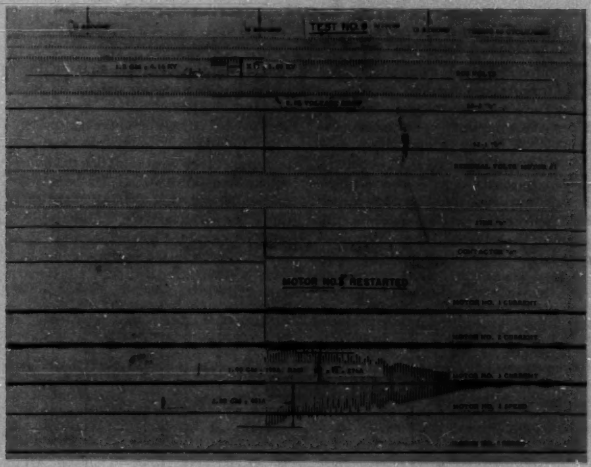


Fig. 11. Oscillographic test 9, motor 3 restarted

time between the maximum breaker clearing time and the minimum upstream reclosing time.

The oscillograph of Figs. 8 through 11 indicated a clearing time longer than was hoped for. Consequently, thought was given to setting the *81* relay up to 59.5 cps. Utility frequency recordings over the past year indicated that such a setting would not result in excessive nuisance tripping due to utility short-time frequency excursions. The relay was recalibrated to 59.5 cps, where it remained during a winter of many severe electrical storms. Experience thus gained indicated that winter disturbances to the utility system shut

Fig. 12. Close-up underfrequency operation, oscillographic test 9

the catalytic reformer down due to frequency excursions often enough to warrant setting the *81* relay back down to a lower value. The new set point after review of the utility frequency recordings over the winter was determined as 585 cps. Thus, such frequency excursions and how often they occur are an important factor in determining the minimum source reclosing time; for this system it is 45 cycles.

Parenthetically, experience has shown that the equipment has repeatedly protected the process. The operators have related that on one black and stormy night filled with lightning the equipment went through the sequence of events pictured in Figs. 8 through 11 thirteen times, keeping the catalytic reformer on stream.

This experience in gaining confidence in the adequacy of the underfrequency relay in properly protecting these motors was time-consuming and expensive. Since the data presented cannot be extrapolated to other systems, but is applicable only to this system, a need is felt for an analytical technique permitting the confident prediction of a system minimum rate of change of frequency. Subsequent to the taking of these data, this need was answered by a confirming analytical technique presented by Morton and Shepherd.

The equation found in this paper for the rate of change of frequency is

$$\Delta f = \frac{0.972HP}{WK^2N^2}\,10^3 = \text{rate of change of frequency}$$

where HP approximated 60 to 75 (essentially losses); $WK^2 = 9{,}500$ lb-ft^2 $N = 400$ rpm.

These values placed in Morton's equation resulted in a calculated rate of change of frequency between 4 and 5; this, while approximate, lends confidence

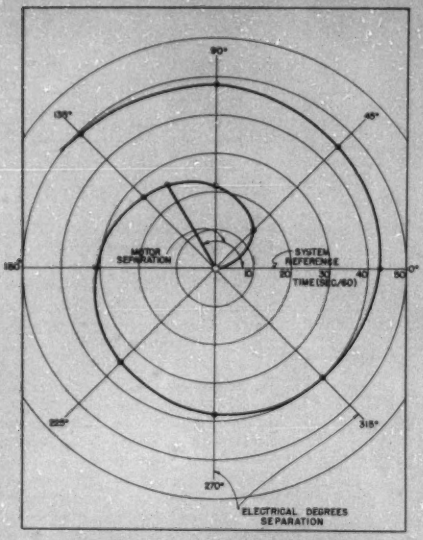


Fig. 13. Electrical separation curve, polar co-ordinates

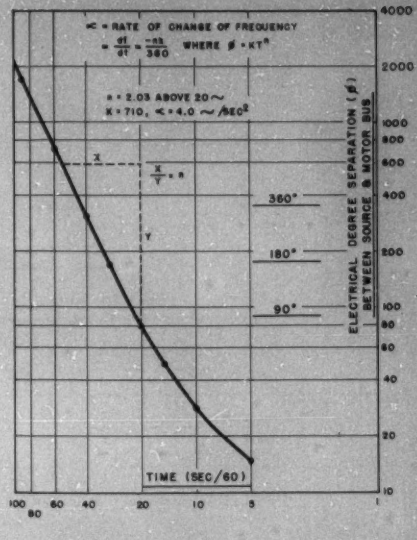


Fig. 14. Electrical separation curve, log-log co-ordinates

to Figs. 8 through 11. The analytical technique and/or oscillograms such as those presented in this paper are the two presently known methods of adequately judging such underfrequency motor protection.

Conclusions

1. High-speed field removal is essential to early residual voltage reduction of a synchronous motor momentarily separated from it's power source.

2. The sequence of relaying under the influence of a multiplicity of conditions is complex. It should be approached with respect and deliberate thought should be devoted to insuring performance. Simplicity, always desirable, should be anticipated only after several attempts. In designing the relaying, generosity proves worth while. The inclusion at the beginning of doubtful functions, ample auxiliary relays, and contacts immensely simplifies the metamorphosis of the system and does so at relatively unimportant cost.

3. Liaison with process people is most important; they must establish the dollar-versus-time evaluation. It would be unwise to put time, money, and effort into a system embodying continuity of an order unwarranted by the process. Conversely, to economize on the relaying, and

Table I. Theoretical Clearing Times*

Times	Cycles			
81 relay setting, cps	58.0	58.5	59.0	59.5
81 relay operating time, at 4 cps/sec	11.5	11.5	11.5	11.5
86-1 relay operating time, cycles	2.0	2.0	2.0	2.0
52-2 circuit-breaker operating time, cycles	6.0	6.0	6.0	6.0
Test equipment operating time, cycles	19.5	19.5	19.5	19.5
Motor system deceleration time to *81* relay setting at 4 cps/sec	30.0	22.5	15.0	7.5
Total clearing time, theoretical, cycles	49.5	42.0	34.5	27.0
Total clearing time, Fig. 8, cycles	41.0			
Suggested minimum source reclosing time, cycles	55	45	40	35

* Estimations except for Fig. 8.

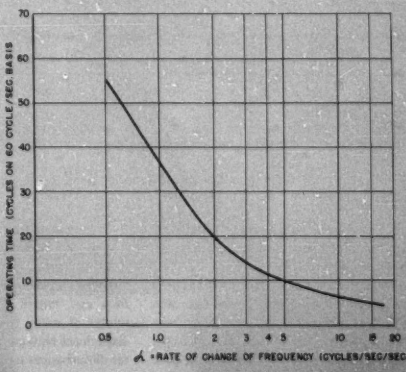


Fig. 15. Operating time of underfrequency relay type CFF13A after system frequency reaches relay setting

by so doing fail to protect the process, would be even more unwise. The "bogey" in connection with the equipment was based on permitting loss of hydrogen pressure supplied by any of the three compressors to a common hydrogen receiver for no more than 10 seconds.

4. Oscillographic records clearly indicate that the equipment meets the process bogey. An even better indication is that the equipment has already faced repeated emergencies and performed far ahead of an operator to save the catalyst.

5. Above a rate of change of frequency of about 10 cps/sec, circuit-breaker clearing time is composed primarily of equipment operating time. Hence, in the design of the system advantage should be taken of every opportunity to reduce the over-all equipment operating time. The best and fastest available devices should be selected and a minimum number of devices used sequentially in time. Over-optimistic estimates should be avoided.

6. However, anything other than the actual minimum rate of change of frequency is of little value in evaluating the merits of the high-speed underfrequency relay and in determining the appropriate minimum time for reclosing the upsteam circuit breakers. By oscillographs, the subject motors and their driven equipment produced a practical minimum rate of change of frequency of 4 cps/sec; this in turn ultimately resulted in a suggested minimum 12-kv circuit-breaker reclosure time of 45 cycles, applicable to this system only.

7. An analytical technique permitting a confident prediction of a system minimum rate of deceleration is helpful in judging the adequacy of underfrequency relaying as synchronous motors protection. This, coupled with oscillographic confirmation, should result in a satisfactory extension of such underfrequency protection to other systems.

8. In arriving at the highest practical underfrequency setting of the underfrequency relay there should be careful consultation with utility records and personnel, with the aim of avoiding operation during short-time underfrequency excursions of the utility. Such excursions do occur but for reasons outside the scope of this paper.

References

1. Transfers of Steam-Electric Generating-Station Auxiliary Buses, D. G. Lewis, W. D. March. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 74, June 1955, pp. 322–34.

2. Transient Protection Conditions in Pipe-Line Stations, E. B. Turner. *Ibid.*, pt. II (*Applications and Industry*), vol. 77, May 1958, pp. 85–90.

3. The Concept of "In Phase" Transfer Applied to Industrial Systems Serving Essential Service Motors, C. C. Young, J. Dunki-Jacobs. *AIEE CP 58-1064 (503)*, (available on request).

4. Transfer Tests on Station Auxiliary Buses, L. E. Backer, Paul Barth, R. A. Hose, D. W. Taylor. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 74, 1955 (Feb. 1956 section), pp. 1441–49.

5. Switchgear with Stored-Energy Mechanism Applied to Steam-Station Auxiliary Transfer Arrangements, P. G. Brown, J. J. Haggerty, D. G. Lewis, E. M. Smith. *Ibid.*, vol. 77, June 1958, pp. 310–18.

6. High Speed Reclosing of Circuit Breakers on Systems Utilizing Synchronous Motor, W. R. Morton, R. V. Shepherd. *AIEE CP 59-891* (available on request).

7. A Transient Stability and Voltage Study for a Modern Oil Refinery Distributions System, C. W. Boice, S. R. Durand, D. Dalasta. *AIEE CP 56-76* (available on request).

8. High Speed Breaker Reclosing Can Put Abnormal Stresses on Your Motors, D. Dalasta, S. Durand. *Power*, New York, N. Y., Design and Equipment Application Section, Feb. 1958, pp. 90–93.

Discussion

F. P. Brightman (General Electric Company, Schenectady, N. Y.): The authors are to be commended for their useful paper. Documented test data is valuable because it confirms the expected behavior of protected and protecting equipment based on calculations, and because it may reveal unsuspected performance characteristics which add to our store of knowledge.

The latter seems to apply here. The authors report that the tests revealed an over-all operating time better than expected, which was an entirely satisfactory conclusion to the tests as far as the particular installation was concerned. However, the fact that the performance was better than could be accounted for by known characteristics of the motors and relays suggests the possibility that some hitherto unknown or unappreciated factors were involved. The purpose of this discussion is to present the results of my attempt to ascertain the reason for the unexpected performance.

I plotted a curve of frequency against time using data obtained from the test 9 oscillogram starting at 5 cycles before the 12-kv circuit breaker opened and ending at 60 cycles after it opened. The resulting curve (Fig. 16) revealed quite unexpected and surprising behavior by the motors. Instead of starting to slow down immediately when their power supply was cut off, the motors apparently speeded up, which is most unorthodox. Then, as shown by the curve, they proceeded to decelerate in an oscillatory fashion, i.e., they alternately slowed down and speeded up in a rhythmic pattern. It is impossible to vouch for detailed accuracy of the oscillations shown in the curve because a difference of 1/10 millimeter in scaling the oscillogram made a change of almost a cycle in the calculated frequency. However, the calculations showed a frequency of 60 cycles for the 5-cycle period prior to opening of the 12-kv circuit breaker, and some form of oscillatory pattern showed up regardless of the method of calculating the data, which leads me to believe that the oscillations were real and not due to errors in scaling the oscillogram.

The data for the frequency-time curve were obtained from a glossy silver print of Fig. 12, which is an enlargement of a portion of the test 9 oscillogram. As a means of compensating for nonuniform movement of the film in the oscillogram (evidenced by unequal lengths of the timing-trace cycles), I compared the lengths of 3 cycles of motor 1 and the timing traces on the oscillogram, and multiplied their ratio by 60 cycles. The use of a 3-cycle span, consisting of the cycle involved and the two adjacent ones, seemed the best compromise between 1. minimizing errors resulting from measurement of the length of the spans, and 2. masking the magnitude of single-cycle oscillations. Curves plotted from data calculated on the basis of measuring the span of 1, 2, 3, and 5 cycles showed the same general pattern. Likewise, the patterns were similar for curves calculated from the O-axis and the upper and lower peaks of the traces.

The curve in Fig. 16 was constructed by plotting the calculated frequency at each succeeding cycle of elapsed time and connecting these points by straight lines. The resulting curve is, of course, unrealistic in that it shows abrupt changes in speed, which are physically impossible because of the mechanical inertia of the equipment, but no greater accuracy was possible with the data available.

Believe it or not, as you wish, the "impossible" frequency-time curve in Fig. 16 accounts for:

1. The unexpectedly short over-all operating time recorded, (41 cycles instead of the authors' calculated 49.5 cycles).
2. The faster and nonlinear angular rate of separation below 20 cycles shown in Figs. 13 and 14.

Fig. 16 shows that the shorter over-all

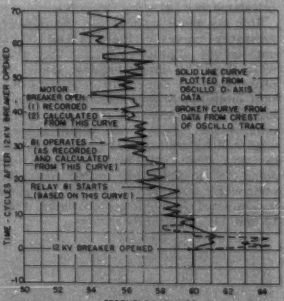


Fig. 16. Frequency-time curve showing behavior of motors based on test 9

operating time resulted because the motors required only 18 cycles to slow down to the 58-cycle setting of the underfrequency relay (device *81*) instead of the calculated 30 cycles. According to Fig. 16 the frequency was constantly fluctuating, and twice during the period from 18 to 30 cycles it even went above the relay's setting of 58 cycles. To determine the relay's operating time for the period, I assumed that the net effect on the relay would be the same as though it had been operating on a constant rate of change of frequency that would have caused the frequency to drop from 58 cycles to the calculated average of 57.3 cycles for the period. Fig. 15 shows that the relay would close its contacts in 12.5 cycles at the 3.5 cps/sec rate of change of frequency calculated on this basis. Adding this to the 18 cycles decelerating time gives a total of 30.5 cycles, which corresponds almost exactly with the recorded operating time of device *81* as shown in Fig. 12. According to the notations of Fig. 12 there was a lapse of 10 cycles from the closing of *81* contacts to the opening of circuit breaker *52-2*. Adding this to the 30.5 cycles gives a total calculated time of 40.5 cycles compared with the recorded 41 cycles reported in the paper.

The curves in Figs. 13 and 14 show that the rate of angular separation during the first 20 cycles was progressively faster as zero time is approached. If the lower end of the separation curve on log-log paper (Fig. 14) is extrapolated at the same slope, the angles of separation at 3 and 4 cycles are 10.7 and 12.8 degrees respectively. The calculated rates of change of frequency during the first 20 cycles based on these extrapolated figures and others taken from the angle of separation curve in Fig. 14 are as follows:

Time, Cycles	Separation Angle, Degrees	Rate of Change, Cycles/Sec/Sec
3	10.7	23.8
4	12.8	16
5	15	12
10	28.5	5.7
20	80	4.0

This pattern of rate of change of frequency corresponds to one that could be calculated from the Fig. 16 curve.

It is my understanding that the angle-of-separation curves in Figs. 13 and 14 did not include allowance for the fact that, when the 12-kv circuit breaker opened the phase angle of the motor, voltage would shift immediately to a position corresponding to that of the motor rotor prior to the interruption. This would mean a backward shift of perhaps 8 degrees for motors carrying about 10% load, as these were doing. Since this shift would have the same effect as motor deceleration, it should be deducted from values shown in Figs. 13 and 14. When this is done,

1. The extrapolation of the Fig. 14 curve would show a negative angle of separation indicating a speed increase rather than decrease in the 0–2 cycle time period.

2. The calculated rate of change of frequency at 20 cycles becomes 3.6 cycles/sec/sec which is essentially the same as the 3.5 value calculated from Fig. 16 for the 18–30 cycle period.

Thus it appears that the curve in Fig. 16 accounts for the performance of the equipment as observed and recorded by the authors. However, we still do not know why the machines behaved that way.

Some of the speculative questions might be:

1. Was the behavior of the machines due to a backward expansion of compressed gas, or to the known fact that synchronous-motor-driven compressors pulsate. In this connection it is interesting to note that the compressor units were in mechanical synchronism during this test, as indicated by the pulsations in the three motor current traces prior to opening of the 12-kv circuit breaker. Consequently, any negative-torque effect would have shown up at its maximum.

2. Was it an isolated incident, or would it happen again to these motors? If so, would it happen regardless of load on the motors?

3. If it was not a freak incident, would it also happen to other motor-driven compressors?

4. Why did the motors slow down so much faster than expected, and would they have behaved proportionately if they had been loaded?

The answers to these questions may have a decided effect on our ability to disconnect compressor motors by means of underfrequency relays fast enough to protect them against out-of-phase re-energization when there is 15–20 cycle reclosing on the power supply.

The authors report that, after these tests were made, the frequency relay was recalibrated to 59.5 cycles to obtain still faster over-all operating time. It was changed again to 58.5 cycles to prevent unnecessary shutdown of the equipment due to power-system frequency fluctuations. That was possible in this case because of ample time for operation of the equipment.

When slowing down of the equipment is not permissible, false operation of the frequency relay can be prevented by using a power relay to monitor it. The contacts of the two relays are connected in series, so that both must operate in order to complete a tripping circuit. With this combination the frequency can fluctuate widely without causing tripping as long as the power relay indicates that the plant is getting power.

Just as a matter of general information, I discovered what was to me a somewhat novel way to scale the oscillogram. First, I made a pinprick at the points on the traces that I wanted to measure, then turned the sheet over and worked on the reverse side. It was much easier to line up the scale on the centers of the pinholes than on the oscillographic traces, especially the somewhat unsymmetrical crests of the waves.

C. L. Phillips and **M. H. Yuen:** Because it is difficult to comment on a field experience without having available the field data being discussed, Mr. Brightman, who had expressed an interest in this subject, was given access to the field data presented in the paper. His discussion is a valuable extension of the interpretation of the data. The authors do not differ with Mr. Brightman's basic views; rather, his effort seemingly confirms the authors' position as follows:

1. More oscillographic data are needed.

2. We do not thoroughly understand what happens to synchronous motor drives within 20 cycles after a power failure.

3. Underfrequency relaying for the purpose of protecting a synchronous motor from a high-speed power reclosure, while apparently the best solution we know today may not permit co-ordination with an upstream high-speed reclosing device.

Questions of the character raised by Mr. Brightman inherently rest on the accuracy of the field data. It should be emphasized that these data were taken by a professional who had spent a lifetime as oscillographer for one of the largest utilities in the country. The authors feel sure that data can be relied upon to be accurate enough to warrant the paper's conclusions, comment, and discussion.

The speculative questions raised regarding the motors' performance are also in the authors' minds; as yet no satisfactory answers have been forthcoming.

Regarding Mr. Brightman's suggestion to utilize a reverse power relay as a blocking relay to utility underfrequency excursions, certain precautions should be mentioned. If it happens that a true tripping underfrequency occurs simultaneously with little or no kilowatt load other than the synchronous motors, then it appears possible that the reverse power relay may fail to remove its blocking function promptly, thus preventing a legitimate trip. Also, if the synchronous motors drive a normally pulsating load, then, at light loading with high power factor the reversing power relay may pulse with the load, negating its blocking function and permitting false trips on utility frequency excursions. Some more positive blocking means is desirable for this critical relay application.

Transistor Amplistat-Regulated Sectional Drive at Southland Paper Mills

A. E. VICKERY
MEMBER AIEE

G. E. SHAAD
MEMBER AIEE

THE HISTORY of the paper industry and the history of paper machine drives is a story of continued progress; progress in improved performance, increased reliability, and operator convenience.

This progress has not been the result of a whim or a slogan. It has come as sheer necessity since the economic foundation of today's paper machines, large or small, is based on the ability continuously to produce a high-quality product. The progressive outlook of the paper industry and a desire for new and better systems has provided the co-operation required between manufacturer and user to make such advances possible.

The progress in machine drives did not take place overnight. It began in the 1920's and has continued to advance through the past two decades.

The early drives of the 1920's were about as crude as can be imagined in the light of today's sophisticated engineering concepts. It must be remembered, however, that the electric sectional drive of those days was only required to provide equal or better performance than the existing mechanical drives that were difficult to maintain, dangerous to personnel, and limited in capacity. Typical of the electric drive of those days was the original drive system installed in Crown Willamette Paper Company, West Lynn, Oreg., in 1920; see Fig. 1. Synchronous motors were connected to the d-c motor shafts through cone pulleys and belts, providing the synchronizing means, while the large slow-speed d-c motors provided the bulk of the driving power.

In the 1920's, this powerful synchronous motor tie was superseded by the mechanical commutator-type rheostat position-regulating systems such as that shown in Fig. 2. The next decade saw the introduction of a position-regulating system utilizing a master selsyn transmitter, belt-driven selsyns on each section, and differential selsyns operating carbon pile resistors in the motor field circuits. By a specialized cam arrangement, the sensitivity of this selsyn system could be maintained essentially constant over a fairly wide range of machine speeds. (The previous systems had reduced sensitivity at reduced speeds.) This selsyn-regulating system furnished throughout the 1930's by the General Electric Company, is shown in Fig. 3.

These various modifications provided considerable improvement in over-all drive performance but by the late 1930's and early 1940's, these regulating systems were reaching the limit of their performance capabilities. It then became necessary to find a system providing faster response with essentially the same regulating accuracy.

To meet this demand for improved performance, several trial installations of high-accuracy speed-regulating systems were made in 1940 and 1941. The performance of these initial installations made it apparent that high-gain speed-regulated systems would provide greatly improved transient performance with more-than-adequate steady-state performance. Fig. 4 shows the electronic amplidyne speed-regulating system furnished by the General Electric Company for the past 10 years. This regulating system consists of a high-accuracy reference system, a d-c tachometer-generator feedback system, an electronic preamplifier, and an amplidyne generator operating in either the motor or the generator field circuits. This high-response speed-regulating system, combined with individual generator power supplies for the various sections, provided excellent performance with great flexibility. From 1946 to 1958, over 100 such machine drives were installed representing more than 1,000 individual regulated sections.

These speed-regulated electronic-amplidyne multiple-generator drives provided very satisfactory performance. The steady-state performance was good and the transient performance shown in Fig. 5 was amply satisfactory for paper-making. From an operator's standpoint, it was convenient, easy to use, fast in response, and generally a quite satisfactory system. In terms of equipment availability and machine lost time attributable to drive performance, the modern drives (installed in the early- and middle-1950's) are averaging well under 2 minutes per day.

Such an operating record would appear to be quite satisfactory but there are always possibilities for improvement. In 1954, the General Electric Company initiated a study program aimed at further reduction of the drive maintenance. Consideration was given to all types of speed-sensing systems, amplifiers, power amplifiers, and generating equipment with due regard for the economics that must always be considered. As a result of this study, a development was initiated to design, test, and evaluate a high-accuracy transistor amplistat speed-regulating system with performance equal to the electronic amplidyne system.

Fig. 1. Sectional drive installed in 1920

Paper 59-813, recommended by the AIEE General Industry Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 23, 1959; made available for printing April 16, 1959.

A. E. VICKERY is with Southland Paper Mills, Lufkin, Tex., and G. E. SHAAD is with the General Electric Company, Schenectady, N. Y.

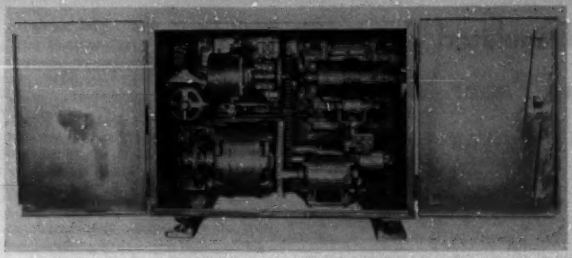


Fig. 2. Rheostat-type position regulator

The attainment of the objective of maintenance without sacrifice of drive performance involved three major points of consideration: (1) circuit design, (2) regulator mechanical design, and (3) over-all control-panel design and system design.

The first consideration represented the major portion of the design and development work. It was necessary to have minimum system drift (due to temperature or other causes), adequate system loop gain for system response and regulation, and high-impedance input circuitry (essentially to impose zero-loading on the high-accuracy d-c tachometer speed-sensing element). Of the many possible methods of approach, the final circuit arrangement was selected on the basis of over-all simplicity and a minimum of components that might require maintenance or replacement. This design is shown in block fashion in Fig. 6. The reference power supply provides an accurately regulated set of bus voltages for the individual section preamplifiers and the master speed control. The master speed reference supply provides a stiff reference voltage (common to all sections) adjustable over the full range from zero to 100% machine speed. Built into this transistorized speed reference system is a linear timing circuit which provides controlled rates of acceleration and deceleration when changing the over-all machine speed.

In each section, the master reference voltage is compared with the section speed in a conventional bridge circuit which permits the use of a simple series rheostat for vernier adjustment (draw) of individual section speeds. This input circuit represents a load on the speed-sensing tachometer generator of about 1 megohm. Any error voltage is amplified and through the power amplistat changes the section generator excitation as required to maintain the section speed.

The drift-free amplification of such low-level high-source impedance signals represented a major design problem. Temperature-controlled ovens' drift could have been controlled by temperature or compensation circuits but the added complications were not desirable. The final design uses a-c amplification of the input signal to convert the initial low-level high-impedance signal into a higher-level lower-impedance signal that no longer represents a problem from a temperature-drift standpoint. A conventional 60-cycle vibrator-chopper is used to convert the d-c input to a-c and demodulate the output of the a-c amplifier to a d-c power output. Similar choppers are used throughout the paper industry in the process instrumentation systems and their record of operational life is exceptionally high.

The remainder of the amplifier system is conventional transistor d-c amplification up to a sufficiently high power level to drive the 400-cycle amplistat power amplifier. It is necessary, of course, to provide a very flexible lead-lag notch network system in the amplifier so that the over-all system performance can readily be matched to the inertia and characteristics of the driven section. In addition, a rate circuit in each section preamplifier provides a controlled rate of acceleration when the section is started.

The basic design was first tested in our own factory in bread-board construction and following these tests a full-sized panel suitable for field testing was constructed. Through the co-operation of a major mill, this trial panel was operated for several months on both high-inertia and low-inertia sections of a high-speed paper machine. From observations of the system performance under conditions of load, temperature, voltage, frequency, etc., a final design was evolved that would meet the operating performance requirements of today's machines and provide the industry with reduced maintenance and freedom from lost time (that can be expected from elements having essentially unlimited life).

The design of the circuit is important but the best of circuits can be inadequate

Fig. 3. Block diagram of electronic amplidyne sectional drive

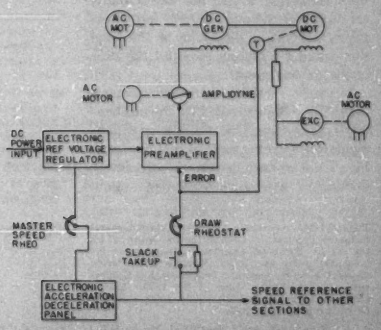


Fig. 4. Carbon pile position-regulator control panel

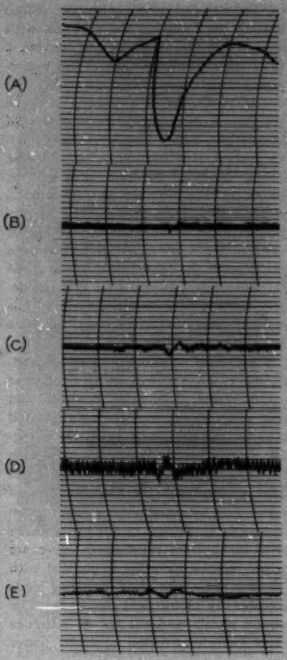


Fig 5 (left).. Speed records resulting from frequency disturbance in an electronic amplidyne regulator

A—Power system frequency
B—Couch to dual press
C—Dual press to smooth press
D—First dryer to first calender
E—First calender to bottom coater

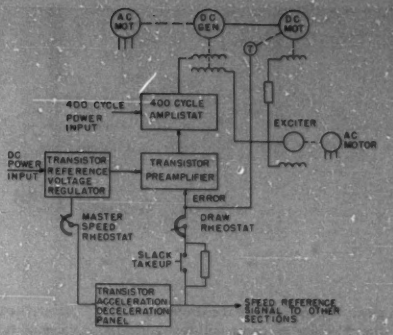


Fig. 6 (right). Block diagram transistor-amplistat sectional drive

if the mechanical design and components are not suitable for the conditions and service involved.

Because corrosion conditions represent a major problem in paper mill design, it was desirable to isolate all small components from the mill atmosphere. It was also desirable to have the potted subsections of such size as to be reasonable in replacement costs. With this in mind, the preamplifier circuit was divided into modules, each a complete subsystem, which were then potted for component protection. Interconnections between modules use high-pressure line contact connectors which have a proved history of maintaining good contact under corrosive conditions.

The complete preamplifier shown in Fig. 7 shows the over-all preamplifier construction.

The amplistat power amplifier is a 400-cycle unit supplied from a separate motor-driven 400-cycle alternator. The selection of 400 cycles was based upon the space and weight of the section amplistats rather than upon the slight increase in speed of response. The mill power supply at 60 cycles is not used for it is desirable completely to isolate the regulating system and the drive performance from the multitude of transient voltage inherent in all mill power systems.

From the regulator to the hermetically sealed control relays and heavy-duty armature components the entire control was designed for the continuous service requirements of the paper industry.

The Southland Paper Mill was conceived and built by its president, Ernest L. Kurth, Sr. It was the first commercial newsprint mill ever built to make newsprint from southern pine. It has grown from one 1,500 fpm (feet per minute) newsprint machine to four newsprint machines totalling 7,800 fpm in a period of 20 years.

The no. 4 paper machine shown in Fig. 8 is a 270-inch Fourdrinier machine of Pusey and Jones manufacture. The drive (motors, gear units, generators, and control) is designed for a speed range of 1,000 to 2,700 fpm on 32-pound basis weight newsprint. The speed is continuously adjustable from a dead stop through 2,700 fpm from the operator's control console at the tending side of the machine.

Considerable over-all planning was

Fig. 7. Transistorized preamplifier panel

Fig. 8. Southland Paper Mills no. 4 machine

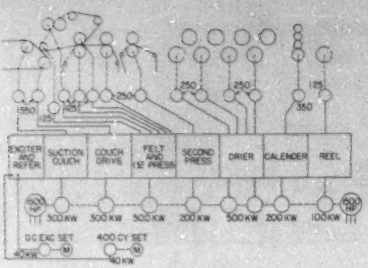


Fig. 9. Southland Paper Mills no. 4 machine drive

given to each item of the drive including the selection of paper-mill-type brushes for the drive motors and generators. Flexibility, reduced maintenance, and reduced down time attributable to electrical repairs and adjustments were important factors in the selection and installation of this drive. Although the drive was only a comparatively small portion of the over-all installation, it could not be overlooked as an important item in reducing maintenance and production down time.

A sectional drive meeting these requirements was the objective in the selection of this drive with transistorized preamplifiers and amplistat (self-saturating magnetic amplifier) power amplifiers. As many moving parts as possible within the regulating loop were eliminated as were other potential sources of trouble such as vacuum tubes and thyratrons. All the conventional and proved regulating systems were considered and finally eliminated in favor of this all-static amplifier.

The amplifier's general physical arrangement is shown in Fig. 9. Its generators were divided into two motor-generator sets, each driven by 1,500-hp (horsepower) synchronous motors. The generators were selected in a minimum number of ratings (three 500 kw, three 300 kw, one 200 kw, and one 100 kw) making for greater flexibility in maintenance and provision for spares. Generator positions on the lineshaft of the motor-generator set are in the same sequence as their respective paper machine sections. The second press generator is mounted on the dry-end motor-generator set to balance the loads on the two sets with a provision for mounting and driving a spare generator from the wet-end motor-generator set.

The sectional drive motors, like the generators, were kept to a minimum of three basic sizes. There are three 350-hp motors, six 250-hp motors, and four 125-hp motors in the drive. The motors, shown in Fig. 10, are of conventional splashproof forced ventilated construction with air entering at the back end of each motor and being discharged into the room at the commutator end. *BC46* permanent-magnet-excited tachometer-generators provide the speed signal for the regulating system and the speed and draw indicators on the machine.

The over-all drive control has six speed-regulated sections: couch drive roll, first press, second press, drier, calender, and reel. In addition to the speed-regulated sections, the suction couch drive is regulated to follow the couch drive roll with the amount of load on the suction couch drive determined by the amount of vacuum in the couch roll. The suction pickup roll, wringer roll, top transfer press, and bottom transfer press drives are helper motors operating in parallel with the first press with their operating loads adjusted by motor field rheostats. The section panels, and the master reference panel (which provides regulated power

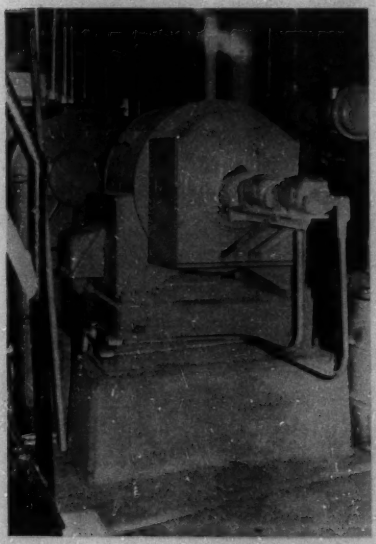


Fig. 10 (left). Section drive motor

Fig. 11 (right). Machine drive control panel

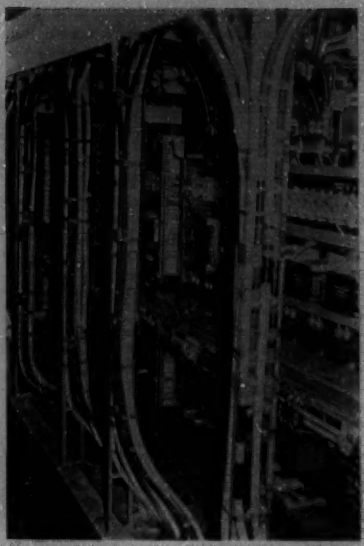


Fig. 12 (left). Back view of drive control panel

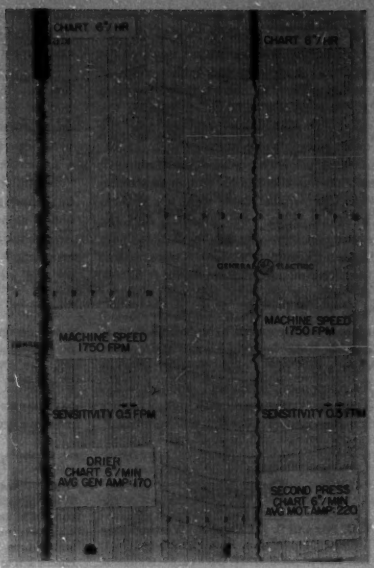


Fig. 14 (right). Speed charts of no. 4 machine

for the preamplifiers, and regulated 400-cycle power for the power amplifiers) are shown in Fig. 11.

A rate amplifier, mounted on the reference panel, provides a definite rate of acceleration and deceleration when the entire speed of the machine is changed. The acceleration and deceleration rates are independently adjustable. The regulated direct current from the reference panel supplies the master speed-adjusting helipot located on the drier operator's console which provides a master reference signal which in turn is fed to the rate amplifier. The rate amplifier output is the reference for all the section speed regulators. This common reference is fed to each of the speed regulators (except the drier regulator) through individual helipots for draw adjustment. (Drier reference is directly from the rate amplifier output.) Slack takeup, where desired, is provided by means of a pushbutton short-circuiting out a resistor in series with the reference circuit.

Fig. 13. Main motor-generator sets and 400-cycle units

The control room in which these panels are installed has a closed air system with air conditioning. The corrosion conditions that had been encountered in other areas of the mill made this construction desirable even though the small regulating components and relays had been sealed to protect them against corrosion and contamination.

The regulator panels mounted on the front of each section panel have no moving parts except for the sealed Stevens-Arnold 6-volt 60-cycle chopper used in the first stage of the regulator. These regulator panels are identical for each section. The units are completely interchangeable with the only adjustment necessary being the setting of the three adjusting rheostats to the same settings as were previously used.

Each section panel is also arranged with essentially identical circuitry permitting maintenance personnel to become readily familiar with the relay sequencing and basic adjustments. The slide-wire resistors for adjustment of crawl speed, regenerative braking current limit, etc., are conveniently mounted on the front of the panel while the larger resistors for motor field adjustments are rack mounted at the top of the panel.

Two modes of operation (at very low or crawl speed) were provided for each section. One uses the transistorized speed regulator and the other uses the amplistat as a voltage regulator. The speed is not adjustable by the operator in either mode but is readily set at the required value by maintenance personnel. To change from transistor speed regulation to amplistat voltage regulation requires the change of one jumper by the electrician. This option applies to slow or crawl speed only. Regulation at operating speeds is always under the control of the transistor amplifier. This amplistat crawl speed option is intended for maintenance purposes only wherein it may be desired to remove one of the transistor amplifier panels for testing during a wire change or other scheduled shutdown. A spare amplifier panel is on hand and can be installed and adjusted in a matter of 20 to 30 minutes and the drive returned to operation. This will allow checking of the removed panel on the bench in a methodical manner without penalizing the machine for excessive electrical down time.

Armored power and control cables were used in making the connections between the section generators, section panels,

section motors, and section operator's panels. Fig. 12 shows a back view of the control panel and the manner in which the incoming and outgoing cables were installed. Fig. 13 shows the two main motor-generator sets together with the 400-cycle motor-generator set and the cable runs to these units.

After all wiring was completed, each section was checked for proper sequential operation, motor fields were set, generator buck fields were adjusted, and the static settings of the regulators were made. Section motors were then operated uncoupled and general regulator performance observed before the motors were coupled to the machine. Once the motors were coupled to their final inertia, the sections were operated throughout their speed range and checked in detail for response, regulation, and stability. These tests were made with high-accuracy fast-response recorders capable of measuring minute errors in the regulating system performance. Fig. 14 showing the drier and second press sections in operation at 1,750 fpm provides a picture of recorder performance. As the chart shows, a minute gear eccentricity exists in the drier section and the 2-cycle-per-second cyclic operation is clearly defined. Full scale on the chart shown in 25 fpm.

A photoelectric recorder (for accurate fast-response recording of section speed) and a standard cathode-ray oscilloscope (for checking the regulator performance) are the basic instruments used in setting up and maintaining the regulating system.

As can be expected in the installation of a new design regulating system, the initial inspection and adjustment involved some changes. These were primarily concerned with added filtering in the regulator panel to eliminate stray voltage pickup that was interfering with regulator performance.

The machine was placed in operation with stock on the wire in October 1958 with no operating problems insofar as the drive was concerned. After several weeks of operation, a 50-ampere silicon rectifier in one of the power amplistats failed and had to be replaced. This failure was located and corrected by the electricians in a minimum of time. After about a month of operation, erratic performance on several sections with a resultant lost time of about 4 hours was traced to loose connections. A subsequent, complete recheck of the panel by the electricians located and tightened many additional loose connections that could cause trouble in future operations.

The experience in these initial months of operation has been very satisfactory. It appears likely that the transistor amplistat regulator will establish a pattern of maintenance and production that will be most satisfactory.

Discussion

E. C. Fox and **S. J. Campbell** (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The use of solid-state devices as regulator amplifiers has proved to be one of the most important steps in recent years to improve reliability and reduce maintenance on paper machine drives. The use of 400-cycle magnetic amplifiers on sectional drives is, of course, not a new development. The Westinghouse Electric Corporation has a total of 12 sectional drives in operation using 400-cycle magnetic amplifiers, the first of which dates back to early 1956. These machines cover a range of products from newsprint to 300-pound board, with operating speeds from 180 to 2,500 fpm.

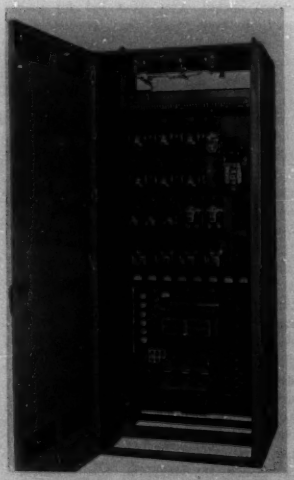


Fig. 13. Typical NEMA I cubicle for one section of a multiple generator sectional drive; 400-cycle magnetic amplifier located below magnetic panel

The paper mentions that the selection of 400 cycles for the magnetic amplifier was based primarily on weight and space considerations. It has been our experience that the trend to higher machine speeds makes necessary more rapid transient response and higher steady-state accuracies. These requirements are more easily met with 400-cycle regulator power supplies as compared to those of 60 cycles. These drives include selenium rectifiers in the regulator circuits, as shown in Fig. 15. Experience has shown that these rectifiers are very reliable. Rectifier aging has had no appreciable effect on regulator operation. Future drives will see an increased use of silicon rectifiers, primarily because of their associated reduction in space requirements. More care must be taken in applying silicon rectifiers than selenium rectifiers because voltage surges are more apt to damage the former.

We should appreciate the authors' comments on three items:

1. Published figures on down time for sectional drives indicate that approximately 32% of electrical down time is chargeable to regulator components and regulator readjustments. Certainly the use of static amplifiers will reduce this figure. However, d-c pilot generators and draw rheostats operating into high-impedance input circuits are still used. We wonder whether any immediate developments are underway to change the nature of the speed cue and draw-adjusting devices to reduce down time chargeable to these devices.

2. The speed charts of Fig. 14 indicate that section speed is being held within 1 fpm at 1,700 fpm. This represents a steady-state accuracy of approximately ±0.03%. Previous papers on tachometer feedback regulators state that the drift of this type of tachometer is guaranteed to be less than 0.03%, and that accuracies as high as 0.04% are required on some machine sections. It is difficult to see how a steady-state accuracy of ±0.3% can be obtained on a prolonged basis when the tachometer drift is of the same order of magnitude. We note that the 60-cycle vibrator chopper and d-c pilot generator have been used in other paper industry applications as elements of integrating amplifiers to offset amplifier drift and to increase the speed regulating accuracy. Is the chopper described in the paper also being used as part of an integrating circuit, as it has on these other applications?

3. The magnetic control and regulators are supplied as open panels, located in a control room with closed air conditioning. This is an excellent example of what appears to be a recent industry trend. We should appreciate the authors' comments on the relative installed cost and operating merits of this system as compared to NEMA (National Electrical Manufacturers Association) I cubicles in a control room supplied with filtered air only.

References

1. Magnetic Amplifier Regulated Drives in the Paper Industry, M. H. Fisher. *TAPPI*, Technical Association of the Pulp and Paper Industry, Easton, Pa., vol. 38, Sept. 1955, pp. 513-22.

2. Integrating Amplifiers Broaden Field for Application of Electric Drives, C. D. Beck. *Electrical Engineering*, vol. 74, no. 10, Oct. 1955, pp. 901-04.

3. Discussion by C. D. Beck, of Control of Speed and Cut in Continuous Processing, E. O. Anger, D. L. Pettit. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, 1954 (Jan. 1955 section), pp. 491-92.

A. E. Vickery and **G. E. Shaad**: The Fox-Campbell discussion covers several points of

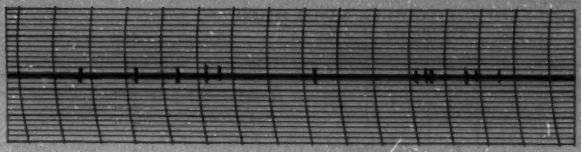


Fig. 16. Tachometer generator accuracy: chart sensitivity 0.01% per small division; chart speed 3 inches per hour

interest that are subject to considerable misunderstanding. They are therefore well worth some additional comment.

The selection of 400-cycle magnetic amplifiers was based upon space and weight, as stated in the paper. A speed-regulated system having adequate preamplifier gain, proper degeneration around the system time constants, and adequate stabilizing components will show little difference between 60-cycle and 400-cycle magnetic amplifiers. Other factors such as roll unbalance, gear backlash, inertia time constants, etc., are limiting on the system response long before the time constants of the magnetic amplifier become a significant factor.

Experience over the past decade has proved beyond any question the reliability of the *BC46* Alnico excited d-c tachometer generators. Care in installation and reasonable preventive maintenance make these units a source of negligible lost time. Improperly installed or maintained, they or any other component can and will cause lost production.

Draw rheostats do not have the 100% reliability record that we should like to obtain. Even so, given proper protection from water and stock, their operating record is quite good. The simplicity of the system has such advantages from a maintenance standpoint that possible rheostat failures become a secondary item.

The d-c tachometer generator has a guaranteed 8-hour drift of 0.03%. This is based upon a 1-megohm, or higher, resistance load in a constant ambient of 20 to 50 degrees centigrade. Actual charts of tachometer performance, given in the 2-hour test chart, show drift to be considerably less than this figure; see Fig. 16. Mill-operating experience has shown that under constant load conditions, the drift and wander (including tachometer generator drift), of a well-designed system will be well within ±0.02%. Many dry end sections require drift and wander performance of this level for satisfactory operation.

The a-c chopper amplifier portion of the preamplifier is not used as an integrating system but merely as a means of providing a-c amplification of the very low level input signal. By the use of a-c amplification at this point the problems associated with temperature drift of transistorized d-c amplifiers are largely eliminated.

The use of open panels with an air-conditioned control room was the result of several factors: It was possible to locate the control room so as to keep unauthorized personnel out of the room. This established, the open panels are convenient to install and to maintain, and they require a minimum control room area. Air conditioning with a closed system was desired from a corrosion standpoint for NEMA I enclosures and normal filtering does not remove the gas contaminants that have caused operating problems in so many mills. There are, of course, methods of removing such gases from the incoming air but in this case air conditioning was the preferred solution.

Input-Output Relationships for Multisampled-Loop Systems

GEORGE G. LENDARIS
ASSOCIATE MEMBER AIEE

ELIAHU I. JURY
MEMBER AIEE

Synopsis: A general gain expression for a multisampled-loop sampled-data system is proposed, and an extended table of systems (Table I) to which this gain expression has been applied is presented.

THE FIRST requisite to the study of sampled-data systems is that an expression for the output as a function of the input and the system variables be found. This is not an easy matter for sampled-data systems of any degree of complexity. The methods that have generally appeared in the literature to date all require a somewhat hit-or-miss type of procedure. The equations for the output are set up, and, in general, equations containing both sampled and unsampled forms of the same variable result. The trick is then to perform analytical sampling (or starring) of one such equation, solve for the sampled variable, and finally search for the right equations in which to substitute. Unfortunately no optimum way of carrying out this procedure has yet been formulated. It would be desirable then to devise a systematic procedure for formulating the output of a sampled-data system in terms of the input and the system variables, as Mason[1,2] has done with continuous systems.

Salzer[3] has presented the technique of replacing the sampler in a sampled-data system by a signal source and defining what he calls the "cut transfer function." He applies his method to single-sampled-loop systems only; this paper will undertake to extend this technique to multiple sampled-loop systems.

This concept of replacing the sampler with a signal source enables one to liken sampled-data systems to multi-input continuous-data systems, and consequently set up the system equations in operational matrix form. Hence a systematic procedure is available for obtaining the output of sampled-data systems.

In general, however, this procedure requires the inversion of matrices, and hence involves a fair amount of computations. Therefore, instead of using this brute force algebraic method, it would be worthwhile to obtain a general expression for the gain of a multisampled-loop system using the kind of topological techniques developed by Mason for continuous systems.

A gain expression is proposed in this paper, and it is shown to work in a large number of systems. The authors have not found any counter examples to the proposed gain expression, and therefore feel that it may be general.

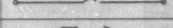
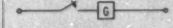
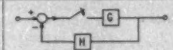
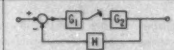
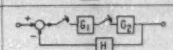
This paper begins by introducing Mason's theorem and proceeds to introduce Salzer's technique of replacing the sampler with a signal source. The ideas of the two methods are considered together and extended to the multisampled-

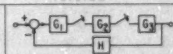
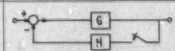
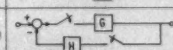
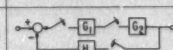
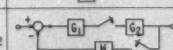
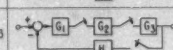
Paper 59-682, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted October 30, 1958; made available for printing March 30, 1959.

GEORGE G. LENDARIS and ELIAHU I. JURY are with the University of California, Berkeley, Calif.

This research was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract no. *AF 18(600)-1321*.

TABLE I INPUT-OUTPUT RELATIONSHIPS FOR SAMPLED SYSTEMS

	SYSTEM	G'_{10}	G'_{1S}	G'_{S0}	G'_{SS}	C
1		0	1	1	0	R^*
2		0	G	1	0	$\overline{RG}^*$
3		0	1	G	0	R^*G
4		0	1	G	$-\overline{GH}^*$	$\frac{R^*G}{1+\overline{GH}^*}$
5		0	G_1	G_2	$-\overline{G_2HG_1}^*$	$\frac{\overline{RG_1}^*G_2}{1+\overline{G_1G_2H}^*}$
6		0	1	G^*	$-G^*H^*$	$\frac{R^*G^*}{1+G^*H^*}$
7		0	1	$G_1^*G_2$	$-G_1^*\overline{G_2H}^*$	$\frac{R^*G_1^*G_2}{1+G_1^*\overline{G_2H}^*}$

	SYSTEM	G'_{10}	G'_{1S}	G'_{S0}	G'_{SS}	C
8		0	G_1	$G_2^*G_3$	$-G_2^*G_3HG_1^*$	$\frac{\overline{RG_1}^*G_2^*G_3}{1+G_2^*G_1G_3H^*}$
9		G	G	$-GH$	$-\overline{GH}^*$	$RG-\frac{\overline{RG}^*GH}{1+\overline{GH}^*}$
10		0	1	G	$-G^*H^*$	$\frac{R^*G}{1+G^*H^*}$
11		0	1	$G_1^*G_2$	$-G_1^*G_2^*H^*$	$\frac{R^*G_1^*G_2}{1+G_1^*G_2^*H^*}$
12		0	G_1	G_2	$-G_2^*\overline{HG_1}^*$	$\frac{\overline{RG_1}^*G_2}{1+G_2^*\overline{HG_1}^*}$
13		0	G_1	$G_2^*G_3$	$-G_2^*G_3^*\overline{HG_1}^*$	$\frac{\overline{RG_1}^*G_2^*G_3}{1+G_2^*G_3^*\overline{HG_1}^*}$

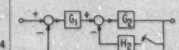
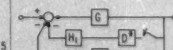
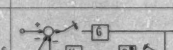
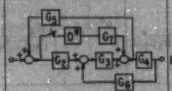
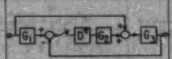
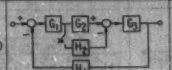
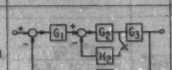
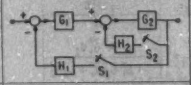
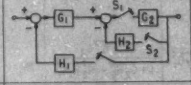
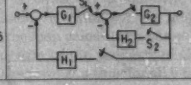
	SYSTEM	G'_{10}	G'_{1S}	G'_{S0}	G'_{SS}	C
14		G_1A $A=\frac{G_2}{1+H_1G_1G_2}$	G_1A	$-H_2A$	$-\overline{H_2A}^*$	$RG_1A-\frac{\overline{RG_1A}^*H_2A}{1+\overline{H_2A}^*}$
15		$\frac{G}{1+GH_2}=B$	B	$-D^*H_1B$	$-D^*\overline{H_1B}^*$	$RB-\frac{\overline{RB}^*D^*H_1B}{1+D^*\overline{H_1B}^*}$
16		0	1	G	$-G^*D^*H_1^*-\overline{GH_2}^*$	$\frac{R^*G}{1+\overline{GH_2}^*+G^*D^*H_1^*}$
17		0	1	G_4E $E=\frac{G_2G_3}{1-G_3G_4G_5}$	$\overline{G_5E}^*$	$\frac{R^*G_4E}{1-\overline{G_5E}^*}$

Table 1 (cont.)

	SYSTEM	G'_{10}	G'_{1S_1}	G'_{S_10}	$G^*_{S_1S_1}$	C
18		$G_2G_3G_4F$ $F=\frac{1}{1-G_5G_6G_7G_8G_9G_6}$	$(1-G_5G_6G_8)F$	$D^*G_1G_4F$	$D^*G_1G_3F^*$	$RG_2G_3G_4F+\frac{\overline{R(1-G_1G_4G_6)F}^*D^*G_2G_4F}{1-D^*\overline{G_1G_3}F^*}$
19		G_3	(G_1-G_3)	$D^*G_2G_3$	$-D^*\overline{G_2}G_3^*$	$RG_3+\frac{\overline{R(G_1-G_3)}^*D^*G_2G_3}{1+D^*\overline{G_2}G_3^*}$
20		$G_1G_2G_3J$ $J=\frac{1}{1+H_1G_1G_2G_3}$	G_1J	$-H_2G_3J$	$+H_2G_3H_1G_1J^*$	$RG_1G_2G_3J-\frac{\overline{RJ}G_1^*H_2G_3J}{1-\overline{H_1H_2}G_1G_3J^*}$
21		$G_1G_2G_3J$ $J=\frac{1}{1+H_1G_1G_2G_3}$	G_1G_2J	$-H_2G_2G_3J$	$-\overline{H_2G_2}J^*$	$RG_1G_2G_3J-\frac{\overline{RG_1G_2J}^*H_2G_2G_3J}{1+\overline{H_2G_2}J^*}$

	SYSTEM	G'_{10}	G'_{1S_1}	G'_{S_10}	$G^*_{S_1S_1}$	G'_{1S_2}	G'_{S_20}	$G^*_{S_1S_2}$	$G^*_{S_2S_1}$	$G^*_{S_2S_1}$	C
22		G_1G_2	G_1G_2	$-H_1G_1G_2$	$-\overline{H_1G_1G_2}^*$	G_1G_2	$-H_2G_2$	$-\overline{H_2G_2}^*$	$-\overline{H_1G_1G_2}^*$	$-\overline{H_2G_2}^*$	$RG_1G_2-\frac{\overline{RG_1G_2}^*H_1G_1G_2+\overline{RG_1G_2}^*H_2G_2}{1+\overline{G_1G_2H_1}^*+\overline{G_2H_2}^*}$
23		0	G_1	G_2	$-G_2^*H_1G_1^*$	0	$H_2^*G_2$	$-H_2^*G_2^*$	G_2^*	$H_2^*G_2^*H_1G_1^*$	$\frac{\overline{RG_1^*}G_2}{1+G_2^*\overline{G_1H_1}^*+G_2^*H_2^*}$
24		0	G_1	G_2	$-G_2^*H_1G_1^*$	0	$-H_2G_2$	$-\overline{H_2G_2}^*$	G_2^*	$\overline{H_2G_2}^*H_1G_1^*$	$\frac{\overline{RG_1^*}(G_2+\overline{H_2G_2}^*G_2-G_2^*H_2G_2)}{1+G_2^*\overline{G_1H_1}^*+\overline{G_2H_2}^*}$
25		0	G_1	G_2	$-G_2^*H_1G_1^*$	0	$H_2^*G_2$	$-G_2^*H_2^*$	G_2^*	$H_2^*G_2^*H_1G_1^*$	$\frac{\overline{RG_1^*}G_2}{1+G_2^*\overline{G_1H_1}^*+G_2^*H_2^*}$

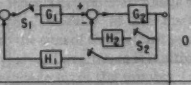
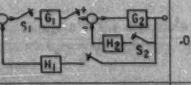
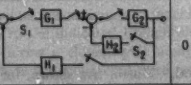
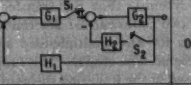
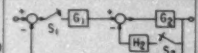
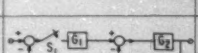
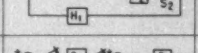
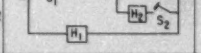
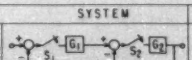
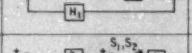
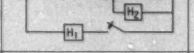
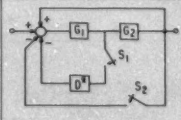
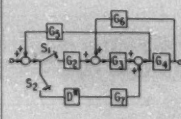
	SYSTEM	G'_{10}	G'_{1S_1}	G'_{S_10}	$G^*_{S_1S_1}$	G'_{1S_2}	G'_{S_20}	$G^*_{S_1S_2}$	$G^*_{S_1S_2}$	$G^*_{S_2S_1}$	C
26		0	1	G_1G_2	$-\overline{G_1G_2}^*H_1^*$	0	$-H_2G_2$	$-\overline{H_2G_2}^*$	$\overline{G_1G_2}^*$	$\overline{H_2G_2}^*H_1^*$	$\frac{R^*G_1G_2+\overline{H_2G_2}^*G_1G_2-\overline{G_1G_2}^*H_2G_2)}{1+\overline{G_1G_2}^*H_1^*+\overline{G_2H_2}^*}$
27		0	1	$G_1^*G_2$	$-G_1^*G_2^*H_1^*$	0	$-H_2G_2$	$-\overline{H_2G_2}^*$	$G_1^*G_2^*$	$\overline{H_2G_2}^*H_1^*$	$\frac{R^*G_1^*(G_2+\overline{H_2G_2}^*G_2-G_2^*H_2G_2)}{1+G_1^*G_2^*H_1^*+\overline{G_2H_2}^*}$
28		0	1	$G_1^*G_2$	$-G_1^*G_2^*H_1^*$	0	$H_2^*G_2$	$-H_2^*G_2^*$	$G_1^*G_2^*$	$H_2^*G_2^*H_1^*$	$\frac{R^*G_1^*G_2}{1+G_1^*G_2^*H_1^*+G_2^*H_2^*}$
29		0	G_1	G_2	$-G_2H_1G_1^*$	0	$-H_2G_2$	$-\overline{H_2G_2}^*$	G_2^*	$\overline{H_2G_2}H_1G_1^*$	$\frac{\overline{RG_1^*}(G_2+\overline{H_2G_2}^*G_2-G_2^*H_2G_2)}{1+G_1G_2H_1^*+\overline{G_2H_2}^*+H_2G_2^*G_2H_1G_1^*-G_2^*H_2G_2H_1G_1^*}$

Table I (cont.)

	SYSTEM	G'_{i0}	G'_{is_1}	$G'_{s_1 0}$	$G^*_{s_1 s_1}$	G'_{is_2}	$G'_{s_2 0}$	$G^*_{s_2 s_2}$	$G^*_{s_1 s_2}$	$G^*_{s_2 s_1}$	C
30		0	1	$G_1 G_2$	$-\overline{G_1 G_2}^*$	0	$-H_2 G_2$	$-\overline{H_2 G_2}^*$	$\overline{G_1 G_2}^*$	$\overline{H_2 G_2}^*$	$\frac{R^*(G_1 G_2 + \overline{H_2 G_2}^* G_1 G_2 - \overline{G_1 G_2}^* H_2 G_2)}{1 + \overline{G_1 G_2}^* + \overline{G_2 H_2}^*}$
31		0	1	$G_1 G_2$	$-\overline{G_1 G_2 H_1}^*$	0	$-H_2 G_2$	$-\overline{H_2 G_2}^*$	$\overline{G_1 G_2}^*$	$\overline{H_2 G_2 H_1}^*$	$\frac{R^*(G_1 G_2 + \overline{H_2 G_2}^* G_1 G_2 - \overline{G_1 G_2}^* H_2 G_2)}{1 + \overline{G_1 G_2 H_1}^* + \overline{G_2 H_2}^* + \overline{G_1 G_2 H_1}^* \overline{H_2 G_2}^* - \overline{G_1 G_2}^* \overline{H_2 G_2 H_1}^*}$
32		0	1	$G_1 G_2$	$-\overline{G_1 G_2 H_1}^*$	0	$-H_2 G_2$	$-\overline{H_2 G_2}^*$	$G_1^* G_2^*$	$\overline{H_2 G_2 H_1}^*$	$\frac{R^* G_1^*(G_2 + \overline{H_2 G_2}^* G_2 - G_2^* H_2 G_2)}{1 + G_1^* \overline{G_2 H_1}^* + \overline{H_2 G_2}^* + G_1^* \overline{G_2 H_1}^* \overline{H_2 G_2}^* - G_1^* G_2^* \overline{H_2 G_2 H_1}^*}$
33		0	1	$G_1 G_2$	$-\overline{G_1 G_2 H_1}^*$	0	$H_2 G_2$	$-H_2^* G_2^*$	$G_1^* G_2^*$	$H_2^* \overline{G_2 H_1}^*$	$\frac{R^* G_1^* G_2}{1 + G_1^* \overline{G_2 H_1}^* + G_2^* H_2^*}$

	SYSTEM	G'_{i0}	G'_{is_1}	$G'_{s_1 0}$	$G^*_{s_1 s_1}$	G'_{is_2}	$G'_{s_2 0}$	$G^*_{s_2 s_2}$	$G^*_{s_1 s_2}$	$G^*_{s_2 s_1}$	C
34		0	1	$G_1^* G_2$	$-G_1^* \overline{G_2 H_1}^*$	0	G_2	$-\overline{G_2 H_2}^*$	$G_1^* \overline{G_2 H_1}^*$	$\overline{G_2 H_1}^*$	$\frac{R^* G_1^* G_2}{1 + G_1^* \overline{G_2 H_1}^* + \overline{G_2 H_2}^*}$
35		0	G_1	G_2	$G_2^* \overline{G_1 H_1}^*$	0	G_2	$\overline{G_2 H_2}^*$	$\overline{G_2 H_2}^*$	$G_2^* \overline{H_1 G_1}^*$	$\frac{R \overline{G_1}^* G_2}{1 + \overline{G_2 H_2}^* + G_2^* \overline{H_1 G_1}^*}$
36		0	1	$G_1^* G_2$	$-G_1^* \overline{G_2 H_1}^*$	0	G_2	$-\overline{G_2 H_2}^*$	$G_1^* \overline{G_2 H_2}^*$	$G_2^* H_1^*$	$\frac{R^* G_1^* G_2}{1 + G_1^* G_2^* H_1^* + \overline{G_2 H_2}^*}$

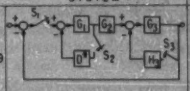
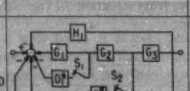
	SYSTEM	G'_{i0}	G'_{is_1}	$G'_{s_1 0}$	$G^*_{s_1 s_1}$	G'_{is_2}	$G'_{s_2 0}$	$G^*_{s_2 s_2}$	$G^*_{s_1 s_2}$	$G^*_{s_2 s_1}$	C
37		$G_2 L$ $L = \frac{G_1}{1+G_1 G_2}$	L	$-D^* G_2 L$	$-D^* L^*$	$G_2 L$	$-G_2 L$	$-\overline{G_2 L}^*$	$-D^* \overline{G_2 L}^*$	$-L^*$	$R G_2 L - \frac{\overline{R L}^* D^* G_2 L + \overline{R G_2 L}^* G_2 L}{1 + D^* L^* + \overline{G_2 L}^*}$
38		0	1	$M = \frac{G_2 G_3 G_4 M}{1 - G_4 G_5}$	$\overline{G_2 G_3 M}^*$	1	$D^* G_1 G_4 M$	$D^* \overline{G_1 G_3 M}^*$	$\overline{G_2 G_3 G_5 M}^*$	$D^* \overline{G_1 G_5 M}^*$	$\frac{R^* G_2 G_3 G_4 M + R^* D^* G_1 G_4 M}{1 - \overline{G_2 G_3 G_5 M}^* - D^* \overline{G_1 G_5 M}^*}$

loop case. From these later considerations the proposed general expression is derived. A list of rules for applying the gain expression and a table of systems to which the expression has been applied are then given.

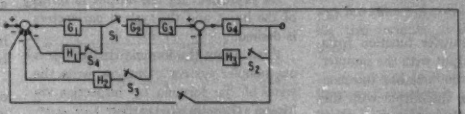
Single-Sampled-Loop Systems

An important background to the derivation of the desired gain expression is the method of one-step reductions of continuous-data linear systems as established by Mason. Mason has shown, for continuous systems, that the transfer gain between two nodes can be determined by inspection for any multiple-loop system from its signal-flow graph. This property is expressed in the following theorem: The

Table I (cont.)

	SYSTEM	G'_D	G'_{1S_1}	G'_{S_1O}	$G^*_{S_1S_1}$	G'_{1S_2}	G'_{S_2O}	$G^*_{S_2S_2}$	G'_{1S_3}	G'_{S_3O}	$G^*_{S_3S_3}$	$G^*_{S_3S_4}$	$G^*_{S_3S_3}$	$G^*_{S_1S_1}$	C
39		0	1	$G_1G_2G_3$	$-G_1G_2G_3^*$	0	$-D^*G_1G_2G_3$	$-D^*G_1^*$	0	$-H_2G_2$	$-\overline{H_3G_2^*}$	G_1^*	$D^*\overline{G_1G_2^*}$	$\overline{H_3G_3}^*$	$\dfrac{R^*G_1G_2G_3(1+H_3G_3^*)-R^*G_1G_2G_3^*H_3G_3}{1+G_1G_2G_3^*+G_1^*D^*+\overline{G_1H_3}^*+D^*G_1^*\overline{H_3G_3}^*}$
40	 $L_1=\dfrac{G_1}{1+G_1G_2G_3H_1}$ $L_2=\dfrac{G_1G_2}{1+G_1G_2G_3H_1}$ $L_3=\dfrac{G_1G_2G_3}{1+G_1G_2G_3H_1}$	L_3	L_1	$-D^*_1L_3$	$-D^*_1L^*_1$	L_2	$-D^*_2L_3$	$-D^*_2L^*_2$	L_3	$-L_3$	$-L^*_3$	$-D^*_1L^*_2$	$-D^*_2L^*_3$	$-L^*_1$	$RL_3=\dfrac{\overline{RL^*_1}D^*_1L_3+\overline{RL^*_2}D^*_2L_3+\overline{RL^*_3}L_3}{1+L^*_1D^*_1+L^*_2D^*_2+L^*_3}$

41



$G'_{1O}=0$ $\quad G^*_{S_1S_2}=\overline{G_2G_3G_4}^*$ $\quad G^*_{S_2S_3}=\overline{H_3G_4}^*G_1^*G_2^*$ $\quad G^*_{S_3S_4}=-\overline{H_2G_1}^*$ $\quad G^*_{S_4S_1}=-\overline{H_1G_1}^*$

$G'_{1S_1}=G_1$ $\quad G'_{1S_2}=0$ $\quad G'_{1S_3}=0$ $\quad G'_{1S_4}=G_1$

$G'_{S_1O}=G_2G_3G_4$ $\quad G'_{S_2O}=-H_3G_4$ $\quad G'_{S_3O}=0$ $\quad G'_{S_4O}=0$

$G^*_{S_1S_1}=\overline{G_2G_3G_4}^*G_1^*$ $\quad G^*_{S_2S_2}=-\overline{H_3G_4}^*$ $\quad G^*_{S_3S_3}=-\overline{H_2G_1}^*G_2^*$ $\quad G^*_{S_4S_4}=-\overline{H_1G_1}^*$

$$C=\frac{\overline{RG_1^*}G_2G_3G_4(1+H_3G_4^*)-\overline{RG_1^*}G_2G_3G_4^*H_3G_4}{1+G_1^*\overline{G_2G_3G_4}^*+\overline{H_3G_4}^*+G_2^*\overline{H_2G_1}^*+\overline{H_1G_1}^*+\overline{H_3G_4}^*H_2G_1^*G_2^*+\overline{H_3G_4}^*\overline{H_1G_1}^*}$$

NOTE: The corrections indicated by Dr. Salzer in his discussion have been incorporated in the table.

transfer gain between any two nodes of a system is equal to a quotient whose denominator is equal to: one minus the sum of the loop gains taken one at a time, plus the sum of the products of the loop gains taken two at a time for all loops that do not touch (have common branch or node), minus the sum of the products of the loop gains taken three at a time for all loops that do not touch, and so on. The numerator of the quotient is equal to the sum of the forward path gains where each of these terms is multiplied by a factor made up of the number *1* and all signed loop gains and products of loop gains that appear in the denominator and do not touch any node or branch found in the forward path.

In this theorem, a feedback loop is a continuous succession of branches, traversed in the indicated branch direction, which forms a closed cycle along which each node is encountered once per cycle. The loop gain of a feedback loop is equal to the product of the gains of the branches forming the loop. A forward path is a continuous succession of branches, traversed in the indicated branch direction from the input node to the output node, along which no node is encountered more than once. The forward path gain is equal to the product of the branch gains along the forward path.

The theorem will be illustrated by calculating the gain of the multiple-loop system shown in Fig. 1. The system has three feedback loops with gains *bc*, *fg*, and *fibj* respectively, and four forward paths with gains *abd*, *efh*, *abjfh*, and *efibd* respectively. Two of the feedback loops (*bc* and *fg*) do not touch. By direct application of the theorem the gain (transfer function) is equal to:

$$C=\frac{abd(1-fg)+efh(1-bc)+abjfh+efibd}{1-bc-fg-fibj+bcfg} \tag{1}$$

It is thus apparent that the direct reduction of even the most complex systems is possible by Mason's method. If this were a sampled-data system in which the signal in each branch were sampled, then all the transfer functions would be sampled-data transfer functions (discrete filters), and all node variables would be sampled functions. The identical procedure just described would constitute a valid one-step reduction, except the notation could be changed by putting an asterisk after every letter in equation 1 and by showing a sampler at the end of each branch in Fig. 1.

The situation is quite different when the system contains both sampled data and continuous data, as would be the case if a sampler were inserted at the beginning of branch *a*, for example. It turns out that the reduction process as discussed could not be directly represented by a transfer function. The problem being that in passing through the sampler in branch *a* the operation is not a multiplication in the frequency domain; rather a summation.

For this reason, one is led toward searching for a technique to handle the sampler. It is at this point that Salzer introduced an artificial signal source in place of the sampler as an aid to analyzing these systems.

To introduce the use of the artificial source, consider the system of Fig. 2. (In this paper it will suffice to show that this artifice works. For a more complete

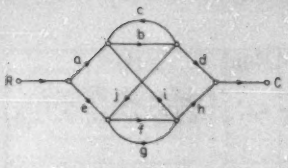


Fig. 1. Flow graph of a multiloop continuous-feedback system

discussion, see reference 3.) If the sampler is opened and a source inserted, call it S^*, and if the transfer functions from input to output, input to sampler, and sampler to output with the sampler open are denoted with a primed notation, then the system output will be given by:

$$C = RG_{ii}' + S^*G_{so}' \tag{2}$$

where G_{ii}' is the transfer function from the input to the output with the sampler open, and G_{so}' is the transfer function from the sampler to the output with the sampler open. If the system had more loops with samplers, then the output would be given by:

$$C = RG_{ii}' + S_1^*G_{s_1o}' + S_2^*G_{s_2o}' + \ldots + S_n^*G_{s_no}' \tag{3}$$

where the prime means that all the samplers are open (except as noted later) when obtaining the given cut-transfer function.

The source S^* is restricted to be equal to the sampled form of the signal entering the sampler.

The signal appearing at the input to the sampler is:

$$S = RG_{ii}' + S^*G_{ss}^* \tag{4}$$

where G_{ss}^* is the transfer function from the output side of the sampler around the loop to the input side of the sampler with the sampler open (and if there were other samplers in the system, they would also be open, except as noted later).

Starring equation 4 the following is obtained:

$$S^* = \overline{RG_{ii}'^*} + S^*G_{ss}^* \tag{5}$$

$$S^* = \frac{\overline{RG_{ii}'^*}}{1 - G_{ss}^*} \tag{6}$$

Substituting equation 6 into equation 2:

$$C = RG_{ii}' + \frac{(RG_{ii}')^*}{1 - G_{ss}^*}G_{so}' \tag{7}$$

Thus a general expression for the output of a one-sampler sampled-data system has been derived. (The discussion to this point has followed closely that of Salzer in reference 3.)

We have at this point managed to 1. bypass the problem presented by the sampler, and 2. find a means by which

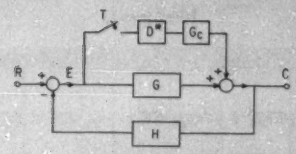


Fig. 2. Single-sampled-loop sampled-data feedback system. D indicates discrete filter (input and output sampled); asterisk replaces sampler that would ordinarily follow box

Mason's methods can be used to find input-output relationships for sampled-data systems.

As an example of this, consider more precisely the system of Fig. 2.

An expression for the output in terms of the input and the system parameters is desired.

First, apply the formula derived for a one-sampler system. One can write the terms of the formula by inspection via Mason's theorem cited earlier:

$$G_{ii}' = \frac{G}{1+GH};\quad G_{is}' = \frac{1}{1+GH};\quad G_{so}' = \frac{D^*G_c}{1+GH}$$

$$G_{ss}^* = -D^*\left(\frac{G_cH}{1+GH}\right)^* \tag{8}$$

Then substituting these terms into equation 7:

$$C = \frac{RC}{1+GH} + \frac{\left(\frac{R}{1+GH}\right)^*}{1+D^*\left(\frac{G_cH}{1+GH}\right)^*}D^*\frac{G_c}{1+GH} \tag{9}$$

Now, for comparison purposes, try to find the output of this system by ordinary algebraic manipulations. The mathematics proceeds as follows:

$$C = EG + E^*D^*G_c \tag{10}$$

$$E = R - CH \tag{11}$$

$$C = RG - CHG + E^*D^*G_c \tag{12}$$

$$C(1+GH) = RG + E^*D^*G_c \tag{13}$$

$$C = \frac{RG}{1+GH} + E^*\frac{D^*G_c}{1+GH} \tag{14}$$

$$E = R - EGH - E^*D^*G_cH \tag{15}$$

$$E(1+GH) = R - E^*D^*G_cH \tag{16}$$

$$E = \frac{R}{1+GH} - E^*D^*\frac{G_cH}{1+GH} \tag{17}$$

$$E^* = \left(\frac{R}{1+GH}\right)^* - E^*D^*\left(\frac{G_cH}{1+GH}\right)^* \tag{18}$$

$$E^* = \frac{\left(\frac{R}{1+GH}\right)^*}{1+D^*\left(\frac{G_cH}{1+GH}\right)^*} \tag{19}$$

And finally, from equation 14 the output is obtained:

$$C = \frac{RG}{1+GH} + \frac{\left(\frac{R}{1+GH}\right)^*}{1+D^*\left(\frac{G_cH}{1+GH}\right)^*}D^*\frac{G_c}{1+GH} \tag{20}$$

It is thus quite apparent that even for this simple system the technique using the formula gives a much more direct answer than does the technique using the algebraic equations. In addition, due to its systematic nature, the formula method is considerably less susceptible to errors.

It will suffice now to say that the method works, and that the answers are obtained in a very straightforward systematic manner. The desired end now is to generalize this method to systems with many sampled loops.

Multisampled-Loop Systems

The extension of this method to multisampled-loop sampled-data systems will be derived along heuristic lines. Following this, some rules will be given, then some examples will be demonstrated. (For a more complete presentation of this subject see the authors' closure.)

In comparing the derivation of the general equations given by Mason for continuous systems, and the derivation of equation 7, it is noted that much similarity exists. By logical inference then, one could investigate the nature of these similarities and make some appropriate extensions.

It has been mentioned before that the output of a system with samplers in n loops is given by the following expression (equation 3):

$$C = RG_{ii}' + S_1^*G_{s_1o}' + S_2^*G_{s_2o}' + \ldots + S_n^*G_{s_no}' \tag{3}$$

The only terms of this expression that have not been defined are the terms S_n^*. This is the quantity that represents the

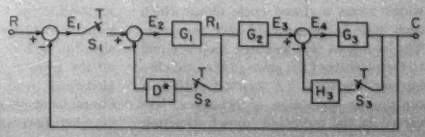


Fig. 3. Multisampled-loop sampled-data feedback system

slave source. This source is obviously dependent upon the input (R) to the system; therefore, it is expected that an expression defining S_n^* would contain a term relating the input and the transference from the input to the switch, hence a term RG'_{0n}. The next question is, how is this term further modified? At this point compare the derivation of equation 6 with those of Mason, and by extension it can be said that the new equation should contain a denominator which would be made up of: one minus the sum of the gains of the sampled loops in the system, plus the sum of the product of all the gains of the sampled loops, taken two at a time, that do not touch each other, etc.

Now all that is left is to define the term sampled-loop gain; this is taken care of very nicely by the term already defined as G^*_{1n1n}. In addition to the terms derived by analogy via Mason's equation, it turns out that for these systems another term is required in the denominator: the product of the self-terms less the product of the cross-terms. That is, the product of the G^*_{1n1n} terms, less the product $G^*_{1n1}G^*_{1n1}G^*_{1n1}\ldots G^*_{1n1}$.

Hence, a general expression for the output of a multisampled-loop sampled-data system with n sampled loops may be

$$C=RG_{10}'+S_1^*G_{20}'+S_2^*G_{30}'+\ldots+S_n^*G_{(n+1)0}' \quad (21)$$

where

$$S_n^*=\frac{\overline{RG_{0n}}^*}{1-\sum_{k=1}^{n}G_{1k1k}^*+\sum P_2-\sum P_3+\ldots+\prod_{k=1}^{n}G_{1k1k}^*-G_{1n1n}^*G_{1n1}^*\ldots G_{1n1}^*}$$

where

$\sum P_2$ = product of loop gains taken two at a time that do not touch each other

$\sum P_3$ = product of loop gains taken three at a time that do not touch each other

This equation then should give the output of any sampled-data system in terms of the input and the system parameters. Apply this equation to a multisampled-loop system and again compare the two techniques for finding the output.

Consider Fig. 3. Again the component terms of the formula can be obtained by inspection. Note that the system contains three sampled loops, two of which do not touch. Therefore, the equation for the slave source will contain an extra product term in the denominator. The terms are as follows:

$$G_{0n}'=0 \quad G_{1n1n}^*=G_1^* \quad G_{2n21}^*=-D^*\overline{G_1G_2G_3}^* \quad G_{3n31}^*=\overline{H_3G_3}^*$$

$$G_{1n1}'=1 \quad G_{1n1}'=0 \quad G_{1n1}'=0$$

$$G_{2n0}'=G_1G_2G_3 \quad G_{2n0}'=-D^*G_1G_2G_3 \quad G_{3n0}'=-H_3G_3$$

$$G_{1n1}^*=-\overline{G_1G_2G_3}^* \quad G_{2n0}^*=-D^*G_1^* \quad G_{3n0}^*=-\overline{H_3G_3}^* \quad (22)$$

The equation for the output follows directly as:

$$C=\frac{R^*G_1G_2G_3(1+\overline{H_3G_3}^*)+R^*G_1G_2G_3^*H_3G_3}{1+G_1G_2G_3^*+G_1^*D^*+G_3H_3^*+D^*G_1^*H_3G_3^*} \quad (23)$$

and

$$C^*=\frac{R^*\overline{G_1G_2G_3}^*}{1+G_1G_2G_3^*+G_1^*D^*+G_3H_3^*+D^*G_1^*H_3G_3^*} \quad (24)$$

Now find the same relationship using standard algebraic techniques. The calculations proceed as follows:

$$C=E_3G_3-C^*H_3G_3 \quad (25)$$

$$=R_1G_2G_3-C^*H_3G_3 \quad (26)$$

$$=E_1G_1G_2G_3-C^*H_3G_3 \quad (27)$$

$$R_1=E_1^*G_1-R_1^*D^*G_1 \quad (28)$$

$$R_1^*=E_1^*G_1^*-R_1^*D^*G_1^* \quad (29)$$

$$R_1^*=\frac{E_1^*G_1^*}{1+D^*G_1^*} \quad (30)$$

$$C=E_1^*G_1G_2G_3-R_1^*D^*G_1G_2G_3-C^*H_3G_3 \quad (31)$$

Let

$$A\triangleq G_1G_2G_3, \; B^*\triangleq G_1^*D^* \quad (32)$$

then

$$C=E_1^*A-\frac{E_1^*B^*}{1+B^*}A-C^*H_3G_3 \quad (33)$$

since

$$E_1^*=R^*-C^* \quad (34)$$

$$C=R^*A-C^*A-R^*A\frac{B^*}{1+B^*}+C^*\frac{B^*}{1+B^*}-C^*H_3G_3 \quad (35)$$

$$C^*=\frac{R^*\left(A^*-A^*\frac{B^*}{1+B^*}\right)}{1+A^*-A^*\frac{B^*}{1+B^*}+\overline{H_3G_3}^*}=\frac{R^*\left(\frac{A^*}{1+B^*}\right)}{1+\frac{A^*}{1+B^*}+H_3G_3^*} \quad (36)$$

$$C^*=\frac{R^*A^*}{1+B^*+A^*+H_3G_3^*+BH_3G_3^*} \quad (37)$$

and finally,

$$C^*=\frac{R^*\overline{G_1G_2G_3}^*}{1+G_1^*D^*+G_1G_2G_3^*+H_3G_3^*+G_1^*D^*H_3G_3^*} \quad (38)$$

The expression for C is obtained by substituting equation 38 into equation 35. The computations are lengthy and will not be included, but the answer is the same as in equation 23.

It is again apparent that the formula gives a much more direct, straightforward solution to the problem. The use of the formula minimizes (in general) the time required to find a solution of this type; however, and more important, due to its systematic nature, it minimizes the chance of error involved in finding the solution.

A few rules will now be given for applying this formula to sampled-data systems.

1. n is equal to the number of separate loops with samplers in them.

2. If a sampler immediately follows a summing junction other than the first, always move it back through the summer before numbering samplers and evaluating the system gains.

3. When evaluating a particular loop gain (primed or starred functions) if the loop has more than one sampler in it, replace the box(es) between each set of samplers with a starred function, and delete the latter sampler(s).

4. When evaluating the cut transfer functions (primed functions) and the G_{3n3n}^* terms, consider all the samplers not in the loop as being open, and use Mason's rule and rule 3 for finding the transmission of the rest of the system.

5. When numbering the samplers, the following must be observed: (a). count only the first sampler in each loop; and (b). the order of numbering must be such that when evaluating the expression $G_{1n1n}^*G_{2n21}^*\ldots G_{2n21}^*$ there are the same number of starred terms as appear in the expression $\prod_{k=1}^{n}G_{1k1k}^*$. (Usually, if a cyclic order of numbering is observed, starting from the major loop around to all the minor loops, this condition is fulfilled.)

Table I showing sampled-data systems analyzed by this method is given for ready reference of the input-output relationships of the given systems and to assist in using this formulation.

The last column in the table is described only by C. The user can do any of three things with this expression: 1. he can consider both sides of the equation as functions of s and thus will have the Laplace transform of the output; 2. he can take the z-transform of both sides of the equation and thus have the output of the system at sampling instants only; and 3.

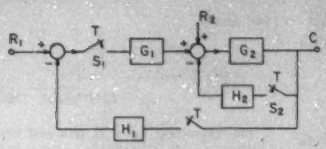


Fig. 4 (left). Multi-input multisampled-loop sampled-data feedback system

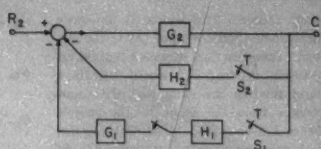


Fig. 5 (right). Fig. 4 considered for R_2 input only

take the modified z-transform of both sides of the equation, and have the continuous output of the system in terms of the z-transform.

All the samplers in the tabled systems are assumed to be in synchronism.

Multiple-Input Systems

This method can easily be extended to multiple-input multiple-sampled-loop sampled-data systems. Since the sampled-data system can be considered as linear (at sampling instants), multiple inputs can be considered as in the continuous case; that is, each input can be considered independently, and by superposition the output can be obtained by adding the components due to each input. For multiple-input multiple-sampled-loop sampled-data systems, one has only; to find the output for each input as described earlier and then add the components to find the total output.

Consider for example, the system of Fig. 4.

The output of this system is desired. By superposition, the output C is the sum of the responses to the two inputs, R_1 and R_2. Considering the system for just R_1, it is noted that the system is the same as no. *26* in Table I. Therefore

$$C_1=\frac{R_1^*(G_1G_2+\overline{H_2G_2}^*G_1G_2-\overline{G_1G_2}^*H_2G_2)}{1+G_1G_2{}^*H_1{}^*+G_2H_2{}^*} \tag{39}$$

Considering the output due to R_2, the system becomes as shown in Fig. 5.

This system is readily solved by comparing it to system *16* in the table. The only term that changes is $G_{22}{}^*$, and it becomes instead $G^*D^*H_1{}^*-G^*H^*_{20}$. With the proper substitutions the output is given by:

$$C_2=\frac{R_2{}^*G_2}{1+G_2{}^*H_2{}^*+G_2{}^*G_1{}^*H_1{}^*} \tag{40}$$

Then the total solution becomes:

$$C=C_1+C_2$$
$$=\frac{R_1^*(G_2G_1+\overline{H_2G_2}^*G_1G_2-\overline{G_1G_2}^*H_2G_2)}{1+\overline{G_1G_2}^*H_1{}^*+G_2H_2{}^*}+\frac{R_2{}^*G_2}{1+G_1{}^*G_2{}^*H_1{}^*+G_2{}^*H_2{}^*} \tag{41}$$

Again it is noted that a quick solution is available using the method herein introduced.

Conclusions

In this paper, an expression giving the output of a multisampled-loop sampled-data system has been heuristically derived. The equation works for all systems of the nature introduced in this paper, and the authors feel that there may not be any difficulties in applying the equation to any linear sampled-data system with all the samplers in synchronism. However, the authors would like to suggest as an area of study that a formal derivation of this expression be found, or at least the derivatiou of an explicit definition of the class of systems for which this formulation will work if it is not absolutely general.

References

1. Feedback Theory—Some Properties of Signal Flow Graphs, S. J. Mason. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Sept. 1953, pp. 1144–56.

2. Feedback Theory—Further Properties of Signal Flow Graphs, S. J. Mason. *Ibid.*, vol. 44, July 1956, pp. 920–26.

3. Signal Flow Reductions in Sampled-Data Systems, J. M. Salzer. *Wescon Convention Record*, Institute of Radio Engineers, pt. 4, 1957, pp. 166–70.

4. Extension of Conventional Techniques in the Design of Sampled-Data Systems, W. K. Linvill, R. W. Sittler. *Convention Record*, Institute of Radio Engineers, pt. I, 1953, pp. 99–104.

5. Design of Sampled-Data Systems by Extension of Conventional Techniques, W. K. Linvill, R. W. Sittler. *Report no. R-221*, Massachusetts Institute of Technology, Cambridge, Mass., July 1957.

6. Extension of Continuous-Data System Design Techniques to Sampled-Data Control Systems, G. W. Johnson, D. P. Lindorff, C. G. A. Nordling. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 74, Sept. 1955, pp. 252–63.

7. Sampled-Data Control Systems (book), E. I. Jury. John Wiley & Sons, Inc., New York, N. Y., 1958, chaps. 1–4.

8. Additions to the Modified z-Transform Method, E. I. Jury. *Wescon Convention Record*, Institute of Radio Engineers, pt. 4, 1957, pp. 136–56.

9. The Geometry of the Zeros of a Polynomial in a Complex Variable, M. Marden. American Mathematical Society, New York, N. Y., 1949, p. 152.

10. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, P. S. Phillips. McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

11. The Analysis of Sampled-Data Systems, J. R. Ragazzini, L. A. Zadeh. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 71, Nov. 1952, p. 225–34.

12. General Synthesis Procedure for Computer Control of Single and Multi-Loop Linear Systems, R. E. Kalman, J. E. Bertram. *Technical Report T-201B*, Columbia University, New York, N. Y., 1957.

13. Conditional Feedback Systems Applied to Stabilizing a Missile in Pitch Attitude, D. R. Katt. *Wescon Convention Record*, Institute of Radio Engineers, 1957, pp. 171–75.

Discussion

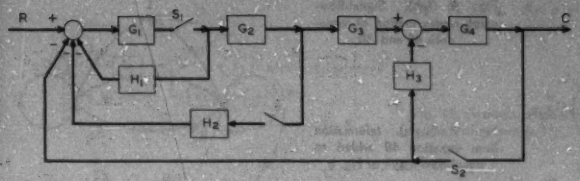
John M. Salzer (Thompson Ramo Wooldridge Inc., Los Angeles, Calif.): The authors have made a notable stab at solving a rather nasty problem of finding the input-output relationship of systems which have several samplers and any number of loops.

The authors reached an intuitive formulation of this solution. Admittedly they do not prove this formulation; they claim, however, that it gives the answer for all the system solutions they attempted.

I would like to emphasize that the solution of such multiloop and multisampled systems is complex, treacherous, and full of pitfalls. Moreover, offering a general solution without proof leaves one full of doubt and uneasiness. The work originally done by the authors probably got close to a solution, but the method and result presented were incorrect and so were many of the specific examples.

Of the specific examples cited (Table I of the paper), the discusser considers the following solutions correct: 1 through 21, 23, 25, 33, 34, 35, 37, 38, and 40. The following solutions were incorrect but contain only what appeared to be slight typographical errors: 22, 28, and 36. The following solutions were conceptually and basically incorrect: 24, 26, 27, 29, 30, 31, 32, 39, and 41.

It might seem peculiar that I claim that the method presented was wrong and yet so many of the solutions turn out to be correct. This is due purely to the fact that in the simpler system configurations, many of the terms of a general solution are zero so that their erroneous absence in the general formula does not influence the answer. As the system configurations become a little more complex, as in the case which this paper was written to solve, the method presented starts producing incorrect answers. It is noted, however, that except for misprints, the authors find the correct denominators in the solution of all examples. This suggests, but does not prove, that the denominator in the authors' generalized equation 21 is correct. Another point to note is that the sampled



$G_{i0} = 0$

$G_{1s_1} = G_1$ $\qquad G_{1s_2} = 0$

$G^*_{2s_1} = -\overline{H_1G_1}^*\,G_2\,\overline{H_2G_1}^*$ $\qquad G^*_{2s_2} = -\overline{H_3G_4}^*$

$G^*_{3s_1} = \overline{G_2G_3G_4}^*$ $\qquad G^*_{3s_2} = -G_1^*$

$G_{3i0} = G_2G_3G_4$ $\qquad G_{3i0} = -G_1^*G_2G_3G_4$

$$C = \frac{\overline{RG_1}^*\,(1+\overline{H_3G_4}^*)\,G_2G_3G_4 - \overline{RG_1}^*\,\overline{G_2G_3G_4}^*\,H_3G_4}{1+\overline{H_1G_1}^* + G_2^*\,\overline{H_2G_1}^* + \overline{H_3G_4}^* + G_1^*\,\overline{G_2G_3G_4}^* + \overline{H_1G_1}^*\,\overline{H_3G_4}^* + G_2^*\,\overline{H_2G_1}^*\,\overline{H_3G_4}^*}$$

Fig. 6. Simplified solution of example 41

form of the system output is correctly given in all examples after the continuous-data part of the numerator is sampled.

It is interesting that the block diagrams in all the examples can be rearranged in a manner requiring only two sampled-data points to interrupt all self-loops. This permits then the solution of all problems here presented by a method adequate for the two-sampled-loop case. For example, the last block diagram 41 is shown to contain five samplers, but it can be rearranged to contain only two sampled-data loops as shown in Fig. 6. The two samplers S_1 and S_4 are adequate to interrupt all loops.

In the process of studying the subject paper, the discusser has arrived at the general equation for a two-sampled-loop system. Using the notation of the article, the equation reads as follows:

$$C = RG_{i0}' + \frac{\overline{RG_{1s_1}}^*(1-G_{2s_2}{}^*)G_{3i0}' + RG_{1s_1}{}^*G_{2s_2}{}^*G_{3i0}' + \overline{RG_{1s_2}}^*(1-G_{2s_1}{}^*)G_{3i0}' + RG_{1s_2}{}^*G_{2s_1}{}^*G_{3i0}'}{1-G_{2s_1}{}^*-G_{2s_2}{}^*-G_{2s_1}{}^*G_{2s_2}{}^*+G_{2s_1}{}^*G_{2s_2}{}^*} \qquad (42)$$

The derivation of equation 42 will be discussed in a forthcoming article. For example 41 of Table I, after rearrangement as shown in Fig. 6, the various quantities are listed and the final formula derived as part of Fig. 6. It is readily visualized that the answer for this example as was given in the article is not correct because it is intuitively clear that H_3 must have an effect on the continuous output of the system. It is one of the interpolating continuous-data transfer functions connecting S_2 to the output.

Equation 42 can be utilized to solve all the other examples.

R. B. Ash, W. H. Kim, and G. M. Kranc (Columbia University, New York, N. Y.): The authors have made an interesting contribution to the theory of sampled-data systems. Their paper extends the topological ideas of Mason to multiloop sampled-data systems. However, the authors' general gain formula, as presented in the paper, does not appear to be valid. First of all, the authors imply in their rule *5* that the application of the formula depends upon a certain technique to be observed in numbering the sampled loops. The authors unfortunately do not reveal this technique. Second, the general expression for $S_n{}^*$, equation 21, is open to question. According to equation 22 the function $S_k{}^*$ is zero. Since the input to sampler S_n, however, is the system output, $S_k{}^* = C^*$, which is certainly not zero.

Equation 21 implies that all one has to do is to place a sampling switch in series with the input in order to make the sampled output identically zero.

Most of the examples which appear in Table I are incorrect. For instance, the output of system 30, as computed algebraically, is

$$C = \frac{R^*G_1G_2 + R^*G_1G_2(G_2H_3)^* - R^*(G_1G_2)^*(G_2H_3)}{1+(G_1G_2)^*+(G_2H_3)^*}$$

Only the first term in the numerator is present in the paper. However, the sampled output:

$$C^* = \frac{R^*(G_1G_2)^*}{1+(G_1G_2)^*+(G_2H_3)^*}$$

agrees with the authors' result.

In almost every example considered by the authors, the actual output contains terms of the form AB^*-A^*B which do not appear in the paper. If the sampled output is taken, these terms drop out, since $(AB^*-A^*B)^*=0$. Thus the sampled output as computed by the authors is invariably correct.

The discussers are preparing a paper on the signal-flow-graph approach to sampled-data systems. A general gain formula, somewhat similar in appearance to that of the authors', will be presented with proof.

George C. Lendaris and Eliahu I. Jury: The authors wish to thank Dr. Salzer for taking the time and effort to study the paper in as much detail as he did and for pointing out that there exists a limitation to equation 21 for the output of an n-sampled-loop system. Also the authors would like to thank Dr. Salzer for indicating the typographical and gross errors that appeared in Table I. These corrections have been incorporated so that the present form of the table should be correct in its entirety.

It may be noted at this point that it was quite fortuitous that the last three entries in the table turned out being $n \leq 2$ systems, for this was not the original intention of the authors. Note also that even though these systems appeared to have more than two loops, the formulation gave the right denominator for the output.

The authors would also like to thank Mr. Ash, Mr. Kim, and Mr. Kranc for their discussion and for their efforts in pointing out the inaccuracies that appear in the paper.

Both discussions indicate that equation 21 is not valid for the general case of the continuous output, but that it may be valid for the general case of the sampled form of the output (at least for $n=2$). Subsequent work by Mr. Lendaris indicates that this latter fact is not correct either. The numerator is not completely general, even for the sampled case.

In his discussion, Dr. Salzer presented a general equation for the continuous output of the $n=2$ system. This equation is correct, for it is a special case of the general solution of the problem which was recently derived.

Messrs. Ash, Kim, and Kranc made a statement in their first paragraph concerning the value of $S_i{}^*$. Their comment implies that they do not fully appreciate the concept of the fictitious signal source as is borne out again by their comment in the succeeding paragraph regarding equation 21.

The authors could continue point by point to discuss the items mentioned by the discussers, however it is much more useful to present instead the subsequent work done in this area by Mr. Lendaris with the purpose of clearing up all the inaccuracies and misconceptions pointed out and perhaps those not yet noted. (who wishes to express thanks to Professor C. A. Desoer for his inspiration and guidance in formulating this material).

For context purposes one might consider the following to be inserted in the paper in place of the section entitled "Multi-sampled-Loop Systems."

Algebraic Method

For extending this approach to the multisampler system, consider Fig. 7 as a

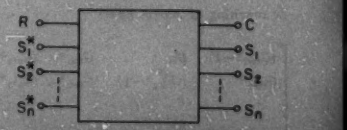


Fig. 7. Multi-input multi-output model of n-sampler sampled-data system

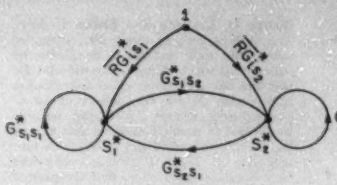


Fig. 8 (left). Signal-flow-graph representation of equations 49 and 50

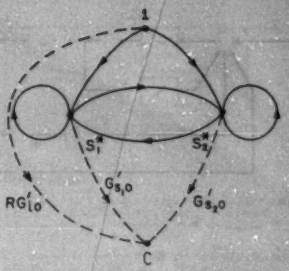


Fig. 9 (right). Information from equation 48 added to signal-flow graph of Fig. 8

model of a sampled-data system. In this model the system has been rearranged such that the input node, the output node, and all of the samplers are outside the box, and the rest of the system is inside the box. The nodes on the right, which will be called output nodes, represent the output terminal of the system, C, and the input terminal of each of the samplers, S_1, S_2, ..., S_n. The nodes on the left, input nodes, represent the system input terminal, R, and the output terminal of each of the samplers, S_1^*, S_2^*, ..., S_n^*. The latter represent the fictitious signal sources introduced earlier in the paper.

The resulting multi-input multi-output system inside the box is a linear time-invariant (because it has no samplers) system, hence there exists a continuous transfer function from each of the input nodes to each of the output nodes. These transfer functions will be denoted by G' with the appropriate subscripts; the first one referring to the input node, the second to the output node.

Keeping these definitions in mind, the following set of equations describe the input-output relationships of the system located inside the box:

$$S_1 = RG_{1O}' + S_1^* G_{S_1S_1}' + \ldots + S_n^* G_{S_nS_1}' \qquad [43(1)]$$

$$\vdots$$

$$S_n = RG_{1n}' + S_1^* G_{S_1S_n}' + \ldots + S_n^* G_{S_nS_n}' \qquad [43(n)]$$

$$C = RG_{1O}' + S_1^* G_{S_1O}' + \ldots + S_n^* G_{S_nO}' \qquad [43(n+1)]$$

Since the sampled signal sources (S_k^*, $k=1, 2, \ldots, n$) are by definition the sampled form of the input to the samplers (S_k, $k=1, 2, \ldots, n$), equations [43(n)] to [43(n)] can be starred and rearranged, yielding:

$$\overline{RG_{1S_1}'^*} = S_1^*(1-G_{S_1S_1}^*) - S_2^* G_{S_2S_1}^* - \ldots - S_n^* G_{S_nS_1}^* \qquad [44(1)]$$

$$\vdots$$

$$\overline{RG_{1n}^*} = -S_1^* G_{S_1S_n}^* - S_2^* G_{S_2S_n}^* - \ldots + S_n^*(1-G_{S_nS_n}^*) \qquad [44(n)]$$

or rewriting in matrix form:

$$\begin{bmatrix} (G_{S_1S_1}^*-1) & G_{S_2S_1}^* & \ldots & G_{S_nS_1}^* \\ G_{S_1S_2}^* & (G_{S_2S_2}^*-1) & \ldots & G_{S_nS_2}^* \\ \vdots & \vdots & & \vdots \\ G_{S_1S_n}^* & G_{S_2S_n}^* & \ldots & (G_{S_nS_n}^*-1) \end{bmatrix} \begin{bmatrix} S_1^* \\ S_2^* \\ \vdots \\ S_n^* \end{bmatrix} = - \begin{bmatrix} \overline{RG_{1S_1}'^*} \\ \overline{RG_{1S_2}'^*} \\ \vdots \\ \overline{RG_{1S_n}'^*} \end{bmatrix} \qquad (45)$$

This set of n equations can be solved for $S_1^*, S_2^*, \ldots, S_n^*$ using standard algebraic techniques. When this is done, the resulting equations can be substituted into equation [43(n+1)] to obtain an expression for C in terms of the input R and the system parameters.

As an example of this method, consider the case where $n=2$. The equations for this system can be written as:

$$S_1 = RG_{1O}' + S_1^* G_{S_1S_1}' + S_2^* G_{S_2S_1}' \qquad (46)$$

$$S_2 = RG_{1O}' + S_1^* G_{S_1S_2}' + S_2^* G_{S_2S_2}' \qquad (47)$$

$$C = RG_{1O}' + S_1^* G_{S_1O}' + S_2^* G_{S_2O}' \qquad (48)$$

Starring and rearranging, equations 46 and 47 become:

$$\overline{RG_{1S_1}'^*} = S_1^*(1-G_{S_1S_1}^*) - S_2^* G_{S_2S_1}^* \qquad (49)$$

$$\overline{RG_{1S_2}'^*} = -S_1^* G_{S_1S_2}^* + S_2^*(1-G_{S_2S_2}^*) \qquad (50)$$

Solving equations 49 and 50 for S_1^* and S_2^*, and substituting these into equation 48, there results:

$$C = RG_{1O}' + \frac{RG_{1S_1}'[G_{S_2O}'(1-G_{S_2S_2}^*) + G_{S_2S_2}^* G_{S_2O}'] + RG_{1S_2}'^*[G_{S_2O}'(1-G_{S_1S_1}^*) + G_{S_2S_1}^* G_{S_1O}']}{1 - G_{S_1S_1}^* - G_{S_2S_2}^* - G_{S_2S_2}^* G_{S_2S_1}^* + G_{S_1S_1}^* G_{S_2S_2}^*} \qquad (51)$$

Equation 51 is the general equation for the output of any sampled-data system containing two samplers (which are in synchronism). This equation is identical to Dr. Salzers' equation 42, and this represents a formal derivation of it.

One could continue in this manner and, using this algebraic method, derive a general expression for the output of an n-sampler system for each n. It is obvious, however, that these general formulas become very lengthy and unwieldy for $n \geq 3$. For this reason, recourse is made to topological methods of finding the output of an n-sampler sampled-data system in terms of the input R and the system parameters.

Topological Method

Consider again the set of equation 45. Instead of solving this set algebraically, one can associate to it a signal-flow graph and then apply Mason's theorem to this signal-flow graph for obtaining the desired relationships.

(The signal-flow-graph techniques of Mason are given in references 1 and 2 of the paper. It should be pointed out that more recent work has been done in this area by Coates.[1] The method he presents is guaranteed to give a minimal term expression for the gain, and for this reason is in many cases superior to Mason's method. His method will not be given in the present discussion, however, so as not to confuse the issue at hand.)

To illustrate this process, consider again the case where $n=2$. From equations 49 and 50 one can readily construct the signal-flow graph shown in Fig. 8. (The input node in this signal-flow graph is labeled with a 1 and the branches leading from it with an RG_{1O}' or $RG_{1S_2}'^*$ term because Mason's theorem gives the transfer gain from one node to the other, and by making the input node variable 1, an expression for the output is obtained directly.) Then adding the information contained in equation 48 the signal-flow graph becomes as shown in Fig. 9. Applying Mason's theorem to find the gain between the nodes labeled 1 and C, one obtains equation 51. This expression, which was constructed directly from the signal-flow graph, again represents the general equation for the output of any sampled-data system containing two samplers (which are in synchronism).

The topological method will not be used, however, to derive the general formulas for the n-sampler sampled-data system; the reason being, as noted earlier, that these general formulas are too lengthy to be practical. Instead, the topological technique will be applied directly to each individual system under consideration. This approach is advantageous because any simplifications arising out of a particular system configuration are borne out in the

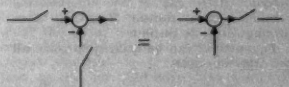


Fig. 10. Example of reducing number of samplers in block diagram

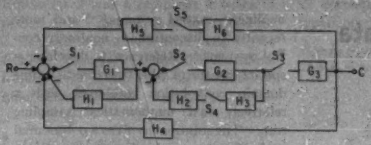


Fig. 11 (left). An $n = 3$ sampled-data system.

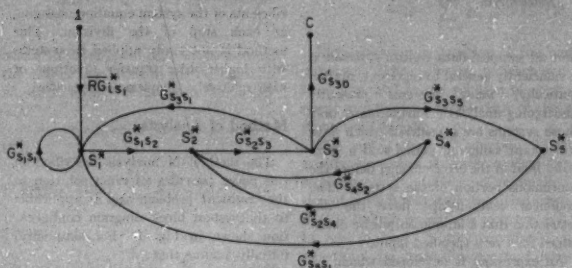


Fig. 12 (below). Signal-flow graph of system in Fig. 11.

signal-flow graph before the calculations are made.

Now that the procedure of constructing a signal-flow graph of the type indicated in Fig. 9 has been justified, one can bypass the process of using the model and the equations used before. Instead, one can go directly from the usual block-diagram form of the sampled system to the signal-flow-graph form. The important thing to notice about the signal-flow-graph representation of the system is that all the node variables are sampled functions, and all the transfer functions are starred transfer functions, hence, as mentioned in the paper, one can apply Mason's theorem to find the desired output. (These comments do not, and need not, apply to the output node nor the branches leading to it.)

To apply the results of the foregoing, use the following procedure:

1. Eliminate all redundant samplers in the original block diagram. See Fig. 10 for an example. This step is not necessary for correct results, but it simplifies the resulting flow graph.

2. Number all samplers in any fashion desired.

3. On a separate sheet of paper draw a node for each sampler indicated in step *2* and label these $S_1^*, S_2^*, \ldots, S_n^*$.

4. Open all samplers in the block diagram.

5. Start from sampler *1* and traverse all possible continuous paths from the output of sampler *1* to the input of all samplers. Where these exist, draw in a branch on the signal-flow graph from node S_1^* to the respective node S_k^* with the direction indicated and the appropriate sampled transfer function. Then start at sampler *2* and repeat the procedure until all samplers are covered.

6. Now add two other nodes: one labeled with a *1* (the input), and one labeled with a *C* (the output). From the input node draw a branch to each of the nodes which has a corresponding continuous path in the block diagram, and label with the appropriate gain (RG_{1k}' to the output, and $\overline{RG_{1k}}^*$ to the other nodes). Then draw a branch from each of the nodes S_k^* which has a corresponding continuous path in the block diagram to the output node *C* and label appropriately (G_{k0}').

7. Apply Mason's theorem between the input node and the output node of the resulting configuration to find the desired output.

As an example of this procedure, consider the system in Fig. 11. The samplers are numbered from *1* to *5*. The signal-flow graph given in Fig. 12 was constructed directly using these rules. Using Mason's theorem on Fig. 12, the output is given by:

$$C = \frac{\overline{RG_{13}}^* G_{31}^* G_{32}^* G_{20}'}{1 - G_{331}^* - G_{332}^* G_{323}^* - G_{323}^* G_{332}^* G_{331}^* - G_{312}^* G_{323}^* G_{331}^* + G_{312}^* G_{332}^* G_{331}^*} \tag{52}$$

where

$$G_{10}' = 1 \qquad G_{331}^* = -\overline{G_1 H_1}^* \quad G_{332}^* = -\overline{G_2 H_1}^*$$

$$G_{323}^* = G_3^* G_{332}^* = \overline{G_2 H_3}^* \qquad G_{12}^* = -H_2^*$$

$$G_{323}^* = G_3^* G_{331}^* = \overline{G_2 H_3}^*$$

$$G_{20}' = G_2 \quad G_{331}^* = -H_1^*$$

This result was obtained with a minimum amount of calculations. If the algebraic method had been used, this whole page would not have sufficed to carry the general formula for the $n = 5$ system.

There are several techniques for minimizing the number of nodes, and hence the computations, in the signal-flow graph, such as, perhaps eliminating nodes S_1^* and S_3^* in Fig. 12 by inspection. This would be done by replacing the appropriate transfer function(s) by a starred transfer function. But due to space limitations these techniques will not be presented in this discussion.

The method presented is completely general, and Mr. Lendaris foresees no difficulties in its application to any *n*-sampler systems no matter how complicated so long as all its samplers are in synchronism.

This concludes the closure to the discussions. We would again like to thank the discussers for indicating the errors that appeared in the paper and for stimulating me to do further work in this area. It is hoped that the short presentation of this material will suffice to clear up any difficulties arising out of the paper, and help to clean up this general topic of signal-flow-graph applications to sampled-data systems.

Reference

1. Flow Graph Solutions of Linear Algebraic Equations, C. L. Coates, *Report no. 58-RL-1997*, General Electric Company, Schenectady, N. Y.; also *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. PGCT-6, no. 2, June 1959, pp. 170–87.

Analysis of Nonlinear Sampled-Data Control Systems—I

E. KINNEN
ASSOCIATE MEMBER AIEE

J. TOU
MEMBER AIEE

AN INTERESTING SIMILARITY exists between the mechanical sampling and zero-order hold function of sampled-data systems, and the mathematical "sampling" which is characteristic of numerical methods of analysis. This paper describes an investigation into this similarity. The results are numerical methods of analysis principally adapted to the transient solution of nonlinear sampled-data control systems. While numerical methods generally provide approximate solutions to nonlinear continuous systems, numerical methods applied to many nonlinear sampled-data systems provide exact solutions. Approximate solutions exist when exact solutions are not possible, i.e., for the more complex system configurations.

The methods developed are useful for the solution of transient problems which arise in systems containing any one of a wide variety of nonlinearities. The procedures are not applicable, however, to problems containing stochastic relationships. The z-transformation is used extensively to provide notational simplicity and to take advantage of existing transform techniques and tables.

In this paper specific systems which can be effectively reduced to the block diagram form appearing in Fig. 1 are considered. The block containing N represents a nonlinear relationship between its input and output terminals, and the block containing G_1 indicates a linear transfer function. Systems containing a nonlinearity between two frequency-sensitive elements are considered in Part II of this report.[1] While the block diagram used in this paper certainly cannot represent all sampled-data control systems, it is sufficiently general to represent, mathematically, many systems presently undergoing analysis. Furthermore, only those systems are considered which have a constant sampling period and a zero-order hold in the error channel before the continuous portion of the system. The nonlinear relationships have the one restriction that a unique output be determined for every specified input.

An expression is developed which, by repetitive application, permits the sampled error values of a nonlinear sampled-data system to be found in response to any specified input disturbance. The step, impulse, and ramp input functions, however, have simplifying advantages. Although the results are easily adapted to machine programming, the specific intent of this development is to provide a practical procedure not dependent on a machine solution.

Known techniques for analyzing nonlinear sampled-data control systems are limited essentially to three: the describing function method, the phase-plane method, and numerical methods. The describing function method, when applied to a system containing a nonlinearity and a sampler, is generally inadequate.[2] Basically, the sampling phenomenon is replaced by a dead time, a procedure which becomes less descriptive as the sampling period approaches the speed of response of the linear portion of the system. Methods using the phase plane have more pertinent application but are limited, practically, to systems with a second-order transfer function. The methods of phase-plane analysis, where possible and sufficiently complete to predict transient phenomena, may also require relatively complex techniques which vary with the type of nonlinearity considered, etc.[3-4]

The principle of numerical methods of analysis has been adapted to continuous nonlinear control systems,[5-7] and more recently to discontinuous nonlinear control systems.[8] Reference 8 develops a special transform table to estimate a time domain convolution combined with a relatively fast mathematical sampling rate. The solution involves a polynomial long division with the nonlinear coefficients of the system equation changing at each step of the division. This method is not easily applied to systems with higher-order transfer functions or with relatively complex nonlinearities.

Method of Analysis

The method of analysis described in this paper provides an exact solution to the transient problem and is applicable to the system block diagram configuration shown in Fig. 1. For simplicity, initially assume that:

1. $G_1(s)$ is the transfer function of the linear frequency sensitive elements of the system and contains no dead time (transport delay).

2. N is a uniquely defined time-independent functional relationship

$$E_{cn}(t) = N[E_c(t)] \tag{1}$$

i.e., any given input condition determines a unique output, dependent only on the instantaneous input magnitude.

3. The sampler function is adequately and unambiguously described as

$$E_c(t) = \sum_{k=-\infty}^{\infty} (E(kT)[U(t-kT) - U\{t-(k+1)T\}]) \tag{2}$$

where T is the sampling period, and $U(t-kT)$ is a positive unit step function occurring at $t=kT$. The lower limit on the summation would be zero if this equation were applied to a system solution which neglects negative time effects or memory.

The effects of system dead time, and storage states in the nonlinearity, are considered separately in the following sections.

From Fig. 1 the Laplace transform of the error can be written as

$$E(s) = R(s) - G_1(s)E_{cn}(s) \tag{3}$$

The Laplace transform of equation 2 is

$$E_c(s) = \sum_{k=-\infty}^{\infty} \left\{ E(kT)e^{-kTs}\left(\frac{1-e^{-Ts}}{s}\right)\right\} \tag{4}$$

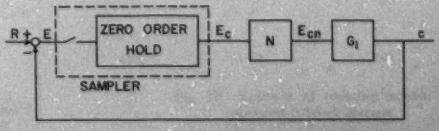


Fig. 1. System configuration for an exact solution

Paper 59-828, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted October 3, 1958; made available for printing May 5, 1959.

E. KINNEN is with the University of Minnesota, Minneapolis, Minn., and J. TOU is with Purdue University, Lafayette, Ind.

The research reported here was supported by the Westinghouse Electric Corporation through a fellowship grant. The authors wish to express their appreciation to Prof. E. M. Sabbagh of Purdue University for suggestions and encouragement.

Equations 2 and 4 represent a series of pulses, each T seconds in duration, spaced T seconds apart and with amplitudes $E(kT)$. Equation 1 states the E_{en} differs from E_e only by the nonlinear relation N, independent of time. Therefore

$$E_{en}(s)=\sum_{k=-\infty}^{\infty}\{N[E(kT)]e^{-kTs}\}\left(\frac{1-e^{-Ts}}{s}\right) \tag{5}$$

where $N[E(kT)]=N[E_k]$ is defined as the output amplitude of the nonlinearity N, as a result of an input amplitude $E(kT)$.

From equations 3 and 5, it follows that

$$R(s)-E(s)=G(s)\sum_{k=-\infty}^{\infty}N[E(kT)]e^{-kTs} \tag{6}$$

where

$$G(s)=G_1(s)\left(\frac{1-e^{-Ts}}{s}\right) \tag{7}$$

Applying the z-transform, equation 6 can be written as

$$R(z)-E(z)=G(z)\sum_{k=-\infty}^{\infty}N[E(kT)]z^{-k} \tag{8}$$

As the information in the z-transform of a time function is contained at sample instants only, equation 8 contains the unknown value of the error function at $t=kT$.

Using a definition of the z-transformation and assuming that $G_1(s)$ is realizable, $G(z)$ can be expanded into a power series of z,

$$G(z)=g_0+g_1z^{-1}+g_2z^{-2}+\ldots g_kz^{-k}\ldots \tag{9}$$

where g_k is the value of the impulse response of $G(s)$, $g(t)$, at $t=kT$. An input function, zero for negative time values, can also be expanded in a power series of z.

$$R(z)=R_0+R_1z^{-1}+R_2z^{-2}+\ldots \tag{10}$$

Equating the unknown error function to a similar power series

$$E(z)=E_0+E_1z^{-1}+E_2z^{-2}+\ldots \tag{11}$$

E_k becomes the value of the unknown error function at $t=kT$. Equation 11 assumes for convenience that the error function is zero for negative time. Substituting equations 9 through 11 into equation 8, it follows that

$$\begin{aligned}&\{(R_0-E_0)+(R_1-E_1)z^{-1}+\\&\quad(R_2-E_2)z^{-2}+\ldots\}\\&=(g_0+g_1z^{-1}+g_2z^{-2}+\ldots)(N[E_0]+\\&\quad N[E_1]z^{-1}+N[E_2]z^{-2}+\ldots)\end{aligned} \tag{12}$$

To satisfy this equation the coefficients of corresponding powers of z must be equal, or

$$\begin{aligned}E_0&=R_0-g_0N[E_0]\\E_1&=R_1-g_0N[E_1]-g_1N[E_0]\\&\vdots\\E_k&=R_k-\sum_{n=0}^{k}g_nN[E_{k-n}]\\&=R_k-\sum_{n=0}^{k}g_{k-n}N[E_n]\end{aligned} \tag{13}$$

Each equation in equation 13 has one unknown. The solution of each unknown is used in the solution of subsequent equations, a characteristic of many numerical methods.

The application of equation 13 is greatly simplified if $g_0=0$, a condition often found in practice. The implication of $g_0=0$ is that sufficient delay exists in $G(s)$ to limit its impulse response, $g(t)$, to zero at $t=0$. Noting equation 7, a minimum of a single time constant delay in $G_1(s)$ will insure this condition. If $g_0=0$, equation 13 reduces to

$$\begin{aligned}E_1&=R_1-g_1N[E_0]\\E_2&=R_2-g_1N[E_1]-g_2N[E_0]\\&\vdots\\E_k&=R_k-\sum_{n=0}^{k}g_nN[E_{k-n}],\ g_0=0\end{aligned} \tag{14}$$

where E_0 is equal to the magnitude of the step disturbance. For each E_k value equation 13 may require the solution of an algebraic equation of the order of the analytic description of N, or a solution by trial and error if N is known graphically. However, if equation 14 is applicable, each E_k is found by determining $N[E_{k-1}]$ and summing a sequence of numerical products. Generally a graphical form of N is best adapted for use with either equation 13 or 14.

Equation 13 can be rearranged and written in matrix form:

$$\|R-E\|=\|G\|\|N[E]\| \tag{15}$$

It is interesting to note that $\|G\|$ in equation 15 is defined as a transmission matrix by Friedland[9] in a discussion of a numerical technique for the analysis of time-varying sampled-data systems.

Another form of equation 14 can be developed by subtracting two consecutive error values, E_k and E_{k-1}. Then, after rearrangement

$$\begin{aligned}R_1&=E_0-a_0N[E_0]\\E_2&=E_1+R_1-R_0-a_0N[E_1]-a_1N[E_0]\\&\vdots\\E_{k+1}&=E_k+R_{k+1}-R_k-a_0N[E_k]-\\&\qquad a_1N[E_{k-1}]-a_2N[E_{k-2}]-\ldots\\&=E_k+R_{k+1}-R_k-\sum_{n=0}^{k}a_nN[E_{k-n}]\end{aligned} \tag{16}$$

where

$$\begin{aligned}a_0&=g_1\\a_1&=g_2-g_1\\&\vdots\\a_k&=g_{k+1}-g_k\end{aligned} \tag{17}$$

The expression of equation 16 has an important advantage over equation 14, as the series a_k often converges rapidly toward zero, whereas the series g_k may converge toward a constant. Consequently the number of terms in equation 16 for an E_k, as k increases, approaches a limit, while those in equation 14 continually increases.

Intersample Analysis

Quite often the information obtained about the system error at sampling instants permits adequate interpolation of the error between sampling instants. When this is not possible or desirable, intersample values may be found at the expense of additional calculations.

Barker[10] (Section 3.4) describes the use of the modified z-transform to obtain intersample information for linear unity feedback systems. A similar method can be used here. Referring again to Fig. 1, a specific error value at any sample time is found by considering a function of all previous error samples $N[E_k]$, and the impulse response of $G(s)$. Specific error values between sample instants consider all previous sampled errors $N[E_k]$, and the delayed impulse response of $G(s)$, the delay limited to a maximum of one sampling period. The modified z-transform, $G(z,m)$ is defined as

$$G(z,m)=z^{-1}\sum_{k=0}^{\infty}g(kT+mT)z^{-k} \tag{18}$$

$0\leq m<1$

The lower limit on the sum is zero, assuming $g(t)$ to be the impulse response of a realizable transfer function.[11] Let g_{mk} be the coefficient of the z^{-k} term in

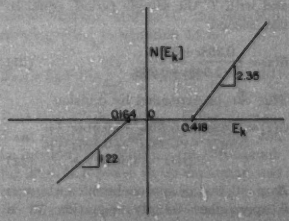


Fig. 2. Simple nonlinearity for Fig. 1

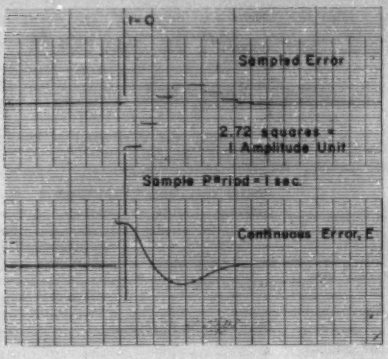


Fig. 3. Error response for a system with a simple nonlinearity

the series expansion of $G(z,m)$, similar to the notation used in equation 9. If the input is expanded as in equation 10, it is readily shown that

$$E_{m0}=0$$

$$E_{mk}=R_k-\sum_{n=1}^{\infty} g_{mn}N[E_{k-n}],\quad k>0 \tag{19}$$

Notice that E_{mk} is not a function of past delayed sampled error values but a function of past sampled error values.

Systems with Simple Nonlinearity, Example I

To illustrate the method of analysis just described, a solution is derived for the system appearing in Fig. 1 with a $G(s)$

$$G(s)=\left(\frac{1-e^{-s}}{s}\right)\left(\frac{1}{s(s+1)}\right) \tag{20}$$

and the asymmetrical simple nonlinearity shown in Fig. 2. A simple nonlinearity is defined as a nonlinear output-input relation dependent only on the instantaneous input magnitude, and independent of any past input values or storage state. Assume a unit positive step function input and all initial error conditions zero. Then, from a standard z-transform table,[13] and with T, the sample period, equal to one second

$$G(z)=\frac{0.368z+0.264}{z^2-1.368z+0.368} \tag{21}$$

This can be expanded by division

$$G(z)=0.368z^{-1}+0.768z^{-2}+0.914z^{-3}+0.967z^{-4}+0.987z^{-5}+0.904z^{-6}+0.997z^{-7}+0.998z^{-8}+\ldots \tag{22}$$

As $g_0=0$, and g_k approaches a constant as k approaches infinity, equation 16 is easier to apply than equation 14. From equation 17 it follows that

$$\begin{aligned}
a_0&=0.368\\
a_1&=0.768-0.368=0.400\\
a_2&=0.146\\
a_3&=0.053\\
a_4&=0.019\\
a_5&=0.007\\
a_6&=0.003\\
a_7&=0.001\\
&\vdots
\end{aligned} \tag{23}$$

$R(z)$, the step function input, is also expanded:

$$R(z)=\frac{z}{z-1}=1+z^{-1}+z^{-2}+\ldots$$

Substituting this and the coefficients of equation 23 into equation 16 yields

$$\begin{aligned}
E_0&=1\\
E_1&=1-0.368N[1]=1-0.368(1.37)=0.497\\
E_2&=0.497-0.368N[0.497]-0.400N[1]\\
&=0.497-0.368(0.186)-0.400(1.37)\\
&=-0.118\\
E_3&=-0.392\\
E_4&=-0.390\\
E_5&=-0.213\\
E_6&=-0.054\\
E_7&=-0.020\\
&\vdots
\end{aligned}$$

The result of an analog computer simulation of this solution for comparison with these calculated values is shown in Fig. 3.

Although a 10-inch slide rule was used for all calculations, the cumulative errors for the calculated values of E_k are within a nominal ± 0.03 experimental reading error for the corresponding values of E_k, through the seventh sampled value.

Systems with Single-Storage Nonlinearity

A single-storage nonlinearity is defined as a nonlinear output-input relation dependent on the instantaneous input magnitude and a storage state created by the one previous input magnitude. If the system in the last example contained a single-storage nonlinearity, such as a hysteresis loop or mechanical backlash, the method of solution would differ only in the evaluation of $N[E_k]$ in the application of equations 13, 14, or 16. $N[E_k]$ would depend on the value of E_{k-1} as well as E_k to establish the input-output uniqueness. For a single-storage nonlinearity, the nonlinear function might be written $N[E_k, E_{k-1}]$.

Consider the idealized saturation curve in Fig. 4. Assume that the system is at rest for negative time, $E_{-k}=0$ and $N[E_{-k}]=0$ for $k>0$, and that the initial error, E_0, is established by a step input occurring at $t=0$. Then, the corresponding nonlinear output must be on the curve *abdoeij* in Fig. 4. Let E_0 be equal to the abscissa value of the point g; $N[E_0]$ is the ordinate value of point g. Then the point g establishes a curve from which the next nonlinear function value of $N[E_1]$ is determined, i.e., *abfgij*. Continuing this procedure for other $N[E_k]$ values, it is apparent that the E_k sequence is just traced along the single-storage nonlinearity according to the basic definition or interpretation of the nonlinearity.

An extension of this modification will permit the inclusion of a multistorage nonlinearity into the system of Fig. 1 which is of the form

$$N[E_k]=N[E_k, E_{k-1}, E_{k-2}, \ldots E_{k-n}] \tag{24}$$

as long as a unique output is determined for each input condition, $(E_k, E_{k-1}, \ldots E_{k-n})$.

To illustrate a problem with a single-storage nonlinearity, consider the $G(s)$ given by equation 20 and the single-storage nonlinearity shown in Fig. 5. Again assume a unit step function input; however, assume an initial constant error value of -0.08 with $N[-0.08]=0$. (This is due to the open-loop drift of the test equipment used to check the solution.)

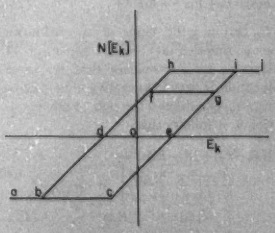


Fig. 4. Single-storage nonlinearity

Combining the coefficients of equation 23 with equation 16 and the nonlinearity given in Fig. 4 yields

$$(E_{-1}=-0.08)$$
$$E_0=-0.08+1.00=0.92$$
$$E_1=0.92-0.368N[0.92]=0.92-0.368(0.644)=0.683$$
$$E_2=0.683-0.368N[0.683]=0.400(0.644)=0.118$$
$$E_3=-0.401$$
$$E_4=-0.823$$
$$E_5=-0.826$$
$$E_6=-0.500$$
$$E_7=-0.056$$
$$\vdots$$

The evaluation of a single-storage nonlinear function can be made easier if the last computed value is placed on a graphical representation of the nonlinearity. Fig. 5 illustrates the E_k values that were plotted for this example calculation.

Fig. 6 shows an analog computer response for this example, easily verifying these computed values. In the method of calculation and the accuracy of the result this example corresponds to the previous illustrative example.

Systems Containing a Transport Delay

The linear portion of the system may include a transport delay which is an even multiple p, of T. Then $G(s)$ has the same form as the right side of equation 7 but multiplied by a term e^{-pTs}; $G(z)$ has the same form as the right side of equation 9 but multiplied by z^{-p}. The equation corresponding to equation 12 is

$$\{(R_1-E_1)+(R_1-E_1)z^{-1}+(R_2-E_2)z^{-2}+\ldots\}=(g_0z^{-p}+g_1z^{-p-1}+g_2z^{-p-2}+\ldots)\times(N[E_0]+N[E_1]z^{-1}+\ldots)$$

where g_k is determined from that part of $G(s)$ which does not include the transport delay.

Equating coefficients of like powers of z, and rearranging, leads to the development of the following equations:

$$E_0=R_0$$
$$E_1=R_1$$
$$\vdots$$
$$E_p=R_p-g_0N[E_0],$$
$$\vdots$$
$$E_k=R_k-\sum_{n=0}^{k-p}g_nN[E_{k-p-n}],\quad k>p \qquad (25)$$

The first p values of the error, all equal to the magnitude of the disturbing function, indicate no system error correction until the initial nonzero error has traversed the transport lag.

The effect of a transport lag which is not an even multiple of T can be investigated by applying the techniques described in the literature.[13,14] Let λT be the transport delay, where λ is a positive number but not an integer. Also let $G_d(s)$ be that portion of $G(s)$ excluding the effect of the transport delay. Define $\lambda=p-\Delta$, where p is a positive integer and $0\leqq\Delta<1$.

$$G(s)=G_d(s)e^{-\lambda Ts}=G_d(s)e^{-pTs}e^{\Delta Ts}$$

As

$$\mathcal{L}^{-1}[G_d(s)e^{\Delta Ts}]=g_d(t+\Delta T)$$

it is possible to define $G_d(z)$ as

$$G_d(z)=\sum_{n=0}^{\infty}g_d(nT+\Delta T)z^{-n}$$

Then, it is possible to write

$$G(z)=G_d(z)z^{-p} \qquad (26)$$

Table II in reference 13, or any modified z-transform table, can be used for the evaluation of $G_d(z)$. If $G_d(z)$ is expanded in a z^{-1} power series, the kth coefficient g_k may be found and substituted into equation 25.

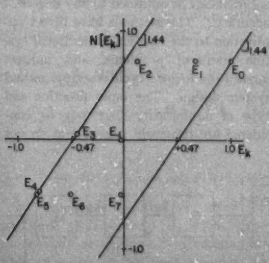


Fig. 5. Single-storage nonlinearity for Fig. 1

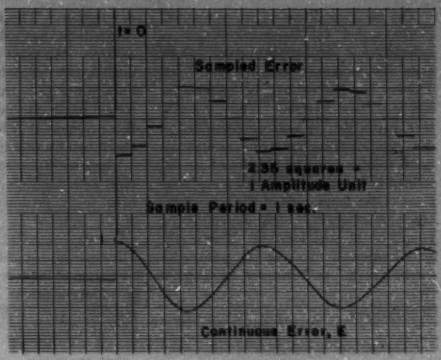


Fig. 6. Error response for a system with a single-storage nonlinearity

Modifications and Extensions

In the examples discussed a simple $G_1(s)$ transfer was used. A more complex $G_1(s)$ only increases the work needed to derive equation 22, the $G(s)$ series expansion. The complexity of subsequent work remains unchanged.

It is apparent from the examples that any form of intricate description of N adds no significant increase in the solution time. Furthermore, as equation 24 may be the effective form of a digital computer command, a generalization of equation 24 will be a nonlinear functional relation not only of all past sampled error values, but also all functions of any past sampled values, etc. Then the computation of any value of the sampled error sequence E_k will depend on the evaluation of the particular numerical program that created the N designated in the block diagram. Although the application to systems containing involved digital programs does not appear attractive, application to dual mode system operation, for example, is straightforward and only as involved as the dual mode switching criterion.

Nonstep input functions are easily considered. If, for example, the input is a unit ramp function beginning at a sample instant, and $g_0=0$, the equation corresponding to equation 14 is

$$E_{k+1}=E_k+T-\sum_{n=0}^{k}g_nN[E_{k-n}] \qquad (27)$$

Equation 17 defines g_k. Other nonconstant inputs, such as a sine function, can be included by obtaining the corresponding $R(s)$ in a power series expansion and combining this with equations 8, 9, and 11. Usually, however, it is not possible to write the kth term in as concise a form as is done in equations 14 or 27.

The methods discussed are not particularly suited to the manipulation of (slowly) time-varying parameters. However, the procedure used by Boxer and Thaler,[7] and Kukel,[8] may be modified for use with these methods. Thus, if a system parameter is not constant but a slowly varying function of time, an approximate solution can be developed as described in this paper but with a readjustment of the magnitude of the time-varying parameters at each increment or sample of the time solution. This would be laborious if the varying parameter changed many terms in the g_k or a_k series.

Conclusions

A significant application of numerical techniques to the analysis of nonlinear sampled-data control systems has been described. The methods developed are relatively easily and quickly applied, requiring only algebraic operations and a z-transform table. The latter is a convenient aid in the initial expansion of the linear portion of the system.

The applications of these methods do not become appreciably more complicated with higher-order system transfer functions or involved nonlinear functional descriptions. This is in direct contrast to conventional methods applied to nonlinear analysis. An exact solution results if the nonlinearity precedes the system frequency-sensitive transfer functions. A second paper considers this problem with the nonlinearity effectively located between two transfer functions.[1]

These methods are well adapted to systems with sample periods approximately equal to the predominant time constants of the system, i.e., for transients significantly controlled by sampler action. They may also quickly provide limited information of limit cycles, following specific initial conditions. This information is generally very difficult to obtain by analytical methods.

References

1. Analysis of Nonlinear Sampled-Data Control Systems—II, E. Kinnen, J. Tou. *AIEE Transactions*, see pp. 390–94 of this issue.

2. Nonlinear Aspects of Sampled-Data Control Systems, R. E. Kalman. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Interscience Publishers, Inc., New York, N. Y., vol. 6, 1957, pp. 273–83.

3. Contactor Servomechanisms Employing Sampled Data, C. K. Chow. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, Mar. 1954, pp. 51–64.

4. Analysis and Design of Relay Type Servomechanisms with Dead Time and Sampling, L. E. Weaver. *Ph.D. Thesis*, Purdue University, Lafayette, Ind., 1958.

5. Sampled-Data Representation of Nonlinear Systems, R. K. Chang. *Ph.D. Thesis*, Purdue University, Lafayette, Ind., 1957.

6. A Step-by-Step Method for Transient Analysis of Feedback Systems with One Nonlinear Element, T. M. Stout. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 75, 1956 (Jan. 1957 section), pp. 378–90.

7. A Simplified Method of Solving Linear and Nonlinear Systems, R. Boxer, S. Thaler. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, Jan. 1956, pp. 89–100.

8. Sampling in Linear and Nonlinear Feedback Control Systems, J. Kukel. *Convention Record*, Institute of Radio Engineers, vol. 5, pt. 4, 1957, pp. 43–56.

9. A Technique for the Analysis of Time-Varying Sampled-Data Systems, B. Friedland. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 75, 1956 (Jan. 1957 section), pp. 407–14.

10. The Pulse Transfer Function and Its Application to Sampling Servosystems (book), R. H. Barker. "Frequency Response", R. Oldenburger, editor. The MacMillan Company, New York, N. Y., 1956, pp. 325–37.

11. Additions to the Modified Z-Transform Method, E. I. Jury. *Convention Record*, Institute of Radio Engineers, vol. 1, pt. 4, 1957, pp. 136–56.

12. Analysis and Synthesis of Sampled-Data Control Systems, E. I. Jury. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, Sept. 1954, pp. 332–46.

13. Sampled-Data Processing Techniques for Feedback Control Systems, A. R. Bergen, J. R. Ragazzini. *Ibid.*, 1954, pp. 236–47.

14. Digital and Sampled-data Control Systems (book), J. Tou. McGraw-Hill Book Company, Inc., New York, N. Y., 1959, pp. 285–89, 588–92.

Analysis of Nonlinear Sampled-Data Control Systems—II

E. KINNEN
ASSOCIATE MEMBER AIEE

J. TOU
MEMBER AIEE

GENERAL METHODS of numerical analysis are developed in a previous paper[1] for nonlinear sampled-data control systems in which the nonlinearity effectively occurs before the linear frequency-sensitive elements in the sampled error channel. These methods are inadequate if the system nonlinearity effectively occurs between two linear frequency-sensitive elements, as shown in Fig. 1. This paper extends the application of numerical analysis to this more complex problem. The resulting methods no longer lead to an exact solution.

If a nonlinearity exists between two frequency-sensitive functions, it is possible to approximate the input to the nonlinearity with a sequence of pulses and then to apply a numerical procedure of analysis. This is done in the following sections. Modified z-transforms and multirate sampler notation are used for brevity in the theoretical development, but only standard z-transform tables are needed in the ultimate applications. The methods are readily adapted to mechanical programming but do not necessarily require a machine solution.

A statement of the general methods which are developed is involved. The application of these methods, however, is not complex because of simplifying approximations that can be made. At the expense of computational labor, the solution can be obtained to any degree of accuracy. The methods have direct application to the solution of nonlinear system transient problems, and indirect application to limit cycle descriptions and system design. The input function and type of nonlinearity that may be considered are relatively general but do not

Paper 59-829, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted October 3, 1958; made available for printing May 5, 1959.

E. Kinnen is with the University of Minnesota, Minneapolis, Minn., and J. Tou is with Purdue University, Lafayette, Ind.

The research reported in this paper was supported by the Westinghouse Electric Corporation through a fellowship grant. The authors wish to express their appreciation to Prof. E. M. Sabbagh, Purdue University, for suggestions and encouragement.

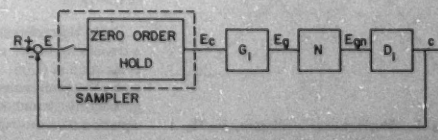


Fig. 1. System configuration for an approximate solution

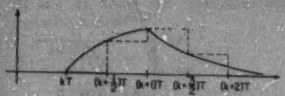
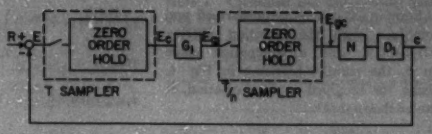


Fig. 2. Pulse response with a straight-line approximation

Fig. 3. Multirate sampler approximation to the system in Fig. 1



include those which are controlled by stochastic processes.

The Problem

Consider the block diagram configuration in Fig. 1; G_1 and D_1 are linear frequency-sensitive transfer functions, while N represents a nonlinear relationship between two signal terminals. Initially assume that: $G_1(s)$ and $D_1(s)$ contain no dead-time delay; N is a uniquely defined time-independent functional relationship of one input; and the sampler function is adequately and unambiguously described as

$$E_s(t)=\sum_{k=-\infty}^{\infty}\{E(kT)[U(t-kT)-U(t-(k+1)T)]\} \quad (1)$$

where T is the sampling period, and $U(t-kT)$ is a positive unit step function occurring at $t=kT$. Modifications of these conditions are discussed following the basic development.

With the use of the notation suggested in Fig. 1, the nonlinear relation is written as

$$E_{gn}(t)=N[E_g(t)] \quad (2)$$

A unique output amplitude at a given time, $E_{gn}(t)$, is dependent on a nonlinear function of a unique input amplitude at the same time, $E_g(t)$, and possible past input magnitudes or storage state. It is also possible to write

$$E_g(t)=E_c(t)*g_1(t) \quad (3)$$

where $g_1(t)$ is the impulse response of $G_1(s)$. It is convenient to combine equations 1 and 3 and insert the result into the functional notation of equation 2.

$$E_{gn}(t)=\sum_{k=-\infty}^{\infty}N\{E_k[g_1^{-1}(t-kT)-g_1^{-1}(t-(k+1)T)]\} \quad (4)$$

where

$$(g_1^{-1}(a)-g_1^{-1}(b))=\int_b^a g_1 dl, \text{ and } E(kT)=E_k$$

The nonlinear function N cannot be combined with D_1 and E_g with linear convolution or superposition techniques. This, of course, is the basic quandary in the analysis of continuous nonlinear systems. A technique is suggested to approximate the summation term of equation 4 by a sequence of zero-slope straight line segments (staircase approximation).

For example, let

$$G(s)=\frac{K_1}{1+\tau s}$$

Then the time function $E_g(t)$ can be shown[3] to have the shape of the smooth curve illustrated in Fig. 2. To make maximum use of the z-transform techniques, a square pulse approximation is made of the time response. These pulses are sketched in Fig. 2 with the use of arbitrarily, three approximating nonzero pulses.

The straight-line or pulse approximation shown in Fig. 2 is simply an effective mathematical sampling operation between G_1 and N. Therefore a second sampler is added to the block diagram, as illustrated in Fig. 3. T is the mechanical sampler period, T/n the mathematical sampler period $n>1$. As a consequence of the added sampler, an approximate solution can be found by modifying the methods of multirate sampled-data system analysis.

Method of Analysis

Following the methods of reference 3, a new block diagram is first constructed (Fig. 4) which is equivalent to the block diagram in Fig. 3. This diagram considers the high-rate sampler as a series of synchronized base-rate samplers, each with a period T, and each placed between a mathematical lead and lag time translation of a multiple of the high-speed sampler period.

When equations 1 and 3 are transformed and combined, $E_g(s)$ becomes

$$E_g(s)=\sum_{k=-\infty}^{\infty}(E(kT)^{-kTs})G(s)(1-e^{-kTs}) \quad (5)$$

where

$$G(s)=\frac{G_1(s)}{s}$$

Referring to Fig. 4, consider only the contribution, $C_p(s)$, of the pth channel to the output $C(s)$. It follows that

$$C_p(s)=N\left[\sum_{k=-\infty}^{\infty}E_d(kT)e^{-kTs}\right]D(s)\times(e^{-pTs/n}-e^{-(p+1)Ts/n}) \quad (6)$$

where

$$D(s)=\frac{D_1(s)}{s} \text{ and } E_d(s)=E_g(s)e^{pTs/n}$$

The N with the bracketed term in this equation, and in those which follow, has the same meaning as the corresponding notation in equation 2, i.e., $N[\ ]$ signifies a unique quantity related in a nonlinear manner to a unique condition represented by the function inside the brackets. Thus, in equation 6

$$N\left[\sum_{k=-\infty}^{\infty}E_d(kT)e^{-kTs}\right]=\sum_{k=-\infty}^{\infty}N(E_d(kT))e^{-kTs}$$

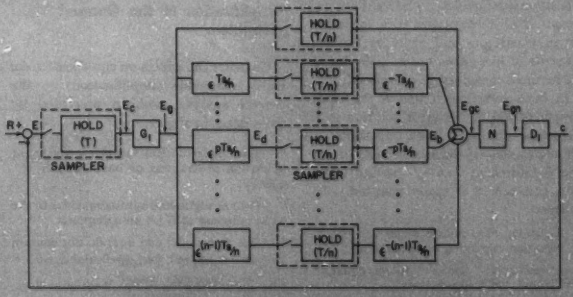


Fig. 4. Multirate sampler approximation to the system in Fig. 1

Equation 6 can be z-transformed and combined with the equation $[R(s)-E(s)=C(s)]$, and a summation made on p, the parallel branches in Fig. 4. Using the modified z-transformation, it can be shown that[2]

$$R(z)-E(z)=\sum_{p=0}^{n-1}\left\{N[E(z)G(z,p/n)(z-1)]\times\left(D(z,1-p/n)-D\left(z,1-\frac{p+1}{n}\right)\right)\right\} \quad (7)$$

Equation 7 can be reduced to a more convenient form:

$$R(z)-E(z)=\sum_{p=0}^{n-1}\left(N\left[E(z)\sum_{u=0}^{\infty}\vartheta_{nu+p}z^{-u}\right]\times\sum_{v=0}^{\infty}\bar{d}_{nv-p}z^{-v}\right) \quad (8)$$

where ϑ_{nu+p} is the $z_n^{-(nu+p)}$ coefficient in the series expansion of the z-transformation

$$Z\left(\frac{G_1(s)}{s}(1-e^{-Ts})\right)\Big|_{z_n=e^{Ts/n}}$$

and $\bar{d}_{nv-p}$ is the $z_n^{-(nv-p)}$ coefficient in the series expansion of

$$Z\left(\frac{D_1(s}{s}(1-e^{-Ts/n})\right)\Big|_{z_n=e^{Ts/n}}$$

The summation on p in the last two equations is a time summation of a sequence of noncoincident pulses, one from each pth channel, cycling consecutively every T seconds. Equations 7 and 8 contain the assumption that the order of $D_1(s)$ is at least s^{-1} as $s\to\infty$, a condition often realized in practice. If this assumption is not valid, the summation in equations 7 and 8 must be from $p=1$ to $p=n-1$ with an added term for the $p=0$ channel.

$R(z)$ and $E(z)$ can be expanded in a power series of known and unknown coefficients respectively, and substituted into equations 7 and 8. The resulting unknown $E(kT)$ values can then be found by an expansion, and equating the coefficients of like powers of z^{-k}, similar to the methods described in reference 1.

In both equations 7 and 8 the basic sample period of the error solution is T seconds; each z in these equations is defined as e^{sT}. If the sampled error solution is desired at $T/n=T'$ second intervals, equation 8 is easily adapted. To do this:

1. Consider $R(z)$ and $E(z)$ on the left of the equation expanded in terms of $e^{sT'}$.
2. Consider z^{-n} and z^{-p} to be equal to e^{-nTs} and e^{-pTs} respectively.
3. Develop ϑ_{nu+p} from

$$Z\left(\frac{D_1(s)}{s}(1-e^{-nTs})\right)\Big|_{z=e^{Ts}}$$

and

$$\bar{d}_{nv-p}\text{ from }Z\left(\frac{D_1(s)}{s}(1-e^{-Ts})\right)\Big|_{z=e^{Ts}}$$

4. In connection with the foregoing, insure that $E(z)$, within the nonlinear brackets, has coefficients of a series expansion which change only to the values of every nth coefficient of the new $E(z)$ solution. Symbolically[3,4]

$$R(z)_n-E(z)_n=\sum_{p=0}^{n-1}\left\{N\left[E(z)\sum_{n=0}^{\infty}\vartheta_{nu+p}z_n^{-u}\right]\times\left(\sum_{v=0}^{\infty}\bar{d}_{nv-p}z_n^{-v}\right)\right\} \quad (9)$$

In equation 9 the subscript n refers to a z-transformation with $z=e^{Ts/n}$.

The term ϑ_{nu+p} can be shown[2] to be the magnitude of the $nu+p$th approximating pulse of the $G_1(s)$ pulse response. In Fig. 2, for example, ϑ_1, ϑ_2, and ϑ_3 are the only nonzero approximating pulses.

Equations 7, through 9 appear to be formidable expressions, but it must be remembered that they are general results. the n represents the number of approximations made for $E_p(t)$ during one base period T. This would be kept to a minimum in practice, say two or three, if possible. The term u is an index on the series expansion of the z-transform of

$$G_1(s)\left(\frac{1-e^{-Ts}}{s}\right)$$

If a sketch, such as Fig. 2, is made, the series $\vartheta_{nu+p}z^{-(nu+p)}$ may be terminated after the coefficients become less than some value, or after a given number of terms. Also, by means of this sketch the computed values of ϑ_{nu+p} could be replaced by any other approximating values.

Simplification of the General Problem

Consider the effects on equation 8 from three reasonable simplifications to the general analysis problem presented by the system in Fig. 1:

1. The mathematical sampler operates at twice the mechanical or base sampler frequency.
2. Four straight-line approximations to the pulse response of $G_1(s)$ are adequate.
3. The terms $g_1(t)$ and $d_1(t)$ do not contain an impulse at $t=0$, i.e., $\beta_0=0$ and $\bar{d}_0=0$.

The approximation in Fig. 2 satisfies these simplifications.

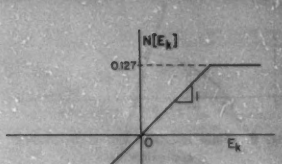


Fig. 5. Simple nonlinearity for illustration

Assume a unit positive step input function. Noting that $n=2$ and $u_{max}=4$, as a result of the first two simplifications, equation 8 can be reduced to

$$[(1-E_0)+(1-E_1)z^{-1}+(1-E_2)z^{-2}+\ldots]=\left\{N[E(z)(\beta_1+\beta_3z^{-1})]\sum_{v=1}^{\infty}\bar{d}_{2v-1}z^{-v}\right\}+\left\{N[E(z)(\beta_2z^{-1})]\sum_{v=1}^{\infty}\bar{d}_{2v}z^{-v}\right\} \quad (10)$$

The terms β_1, β_2, and β_3 are the magnitudes of the three nonzero straight-line approximations to the pulse response of $G_1(s)(1-e^{-Ts})/s$. Solving for the coefficients of the unknown error series, the following equations result:

$$E_0=1$$
$$E_1=E(T)=1-\bar{d}_1N[E_0\beta_1]$$
$$E_2=E(2T)=1-\bar{d}_1\{N[E_1\beta_1]+N[E_0\beta_2]\}-\bar{d}_2N[E_0\beta_1]-\bar{d}_3N[E_0\beta_1]$$
$$E_3=E(3T)=1-\bar{d}_1\{N[E_2\beta_1]+N[E_0\beta_1]\}-\bar{d}_2N[E_1\beta_1]-\bar{d}_3\{N[E_0\beta_2]+N[E_1\beta_1]\}-\bar{d}_4N[E_0\beta_1]-\bar{d}_5N[E_0\beta_1]$$
$$\vdots$$

or

$$E_k=E_{k-1}-\sum_{z=1}^{k}(\bar{d}_{2z-1}-\bar{d}_{2z-2})\times(N[E_{k-z}\beta_1]+N[E_{k-z-1}\beta_1])-\sum_{y=1}^{y=i}(\bar{d}_{2y}-\bar{d}_{2y-2})(N[E_{k-y-1}\beta_1]) \quad (11)$$

In the equation for a general E_k, the terms $(\bar{d}_{2z-1}-\bar{d}_{2z-2})$, and $(\bar{d}_{2y}-\bar{d}_{2y-2})$ usually approach zero rapidly. ϑ and $\bar{d}$ are defined with equation 8.

Illustrative Example

For an example, consider the system in Fig. 1 with

$$G_1(s)=\frac{0.405}{(s+1)^2}\quad D(s)=\frac{2}{s(s+2)}$$

$T=1$ second, $R(s)=1/s$, and the simple N shown in Fig. 5. Let $n=2$, the mathematical sampler operating at twice the mechanical sampler frequency.

The two expansions which determine

$\bar{g}_3$ and $\bar{g}_4$, using the expressions with equation 8, are

$$Z\left(\frac{0.405}{s(s+1)}(1-e^{-s})\right)\Big|_{s=\rho,\omega} = \frac{0.1597}{z}\left(\frac{z+1}{z-0.606}\right) = \bar{g}(z)$$
$$= 0.1597z^{-1} + 0.2565z^{-2} + 0.1552z^{-3} + 0.0941z^{-4} + 0.0572z^{-5} + \ldots \quad (12)$$

$$Z\left(\frac{2}{s^2(s+2)}(1-e^{-0.5s})\right)\Big|_{s=\rho,\omega} = \frac{0.184z+0.132}{z^2-1.368z+0.368} = \bar{d}(z)$$
$$= 0.184z^{-1} + 0.384z^{-2} + 0.457z^{-3} + 0.484z^{-4} + 0.494z^{-5} + 0.497z^{-6} + \ldots \quad (13)$$

Consider only the first three terms of equation 12 ($u_{max}=4$, but $g_0=0$). Then, the three simplifications mentioned in the last section are satisfied and equation 11 is applicable. Note the convergent sequence:

$\bar{d}_1 = 0.184$
$\bar{d}_2 = 0.384$
$\bar{d}_3 - \bar{d}_1 = 0.273$
$\bar{d}_4 - \bar{d}_2 = 0.100$
$\bar{d}_5 - \bar{d}_3 = 0.037$
$\bar{d}_6 - \bar{d}_4 = 0.013$
$\vdots$

Then, it follows that

$E_0 = 1$
$E_1 = E_0 - \bar{d}_1 N[E_0\phi_1] = 1 - 0.184N[0.1597]$
$= 1 - 0.184(0.127) = 0.977$
$E_2 = E_1 - \bar{d}_2[N[E_0\phi_1] + N[E_0\phi_0]] - (\bar{d}_3 - \bar{d}_1)N[E_0\phi_1] - \bar{d}_2 N[E_0\phi_2] = 0.977 - 0.184(0.127 + 0.127) - 0.273(0.127) - 0.284(0.127) = 0.847$
$\vdots$

The results are given in Table I. For comparison and checking, this example was placed on an analog computer with two samplers, the first operating at 1 cycle per second as stated in the problem, and the second operating at 2 cycles per second, simulating $n=2$ in Fig. 4. The computer solution, however, does not limit u_{max} to four, but considers all terms in the $g(s)$ series. The computer results are shown in Fig. 6 with the sampled error values also listed in Table I. To check the approximation caused by including a second sampler and by letting $n=2$, the computer results for the problem without the second sampler are given in Fig. 7. The sampler error values in Table I from the solution show very close correspondence to those from the approximate solutions. The computed results, including the seemingly severe roundoff effects of equation 11, the accuracy of the measured quantities, and the 10-inch slide rule calculations, are within 10% of the maximum error through the seventh sampled value.

Table I. Calculated and Experimental Sampled Error Values for the Numerical Example

Sampled Error	Approximation: Calculated $n=2$, $u=4$	Approximation: Experimental $n=2$, $u=\infty$	Experimental
E_0	1.000	1.00	1.00
E_1	0.977	0.97	0.94
E_2	0.847	0.82	0.77
E_3	0.705	0.68	0.60
E_4	0.480	0.47	0.43
E_5	0.311	0.28	0.24
E_6	0.169	0.10	0.08
E_7	0.053	-0.05	-0.06
		± 0.03	± 0.03

A comparison of the two experimental results suggests that the slowness of the calculated response is due to the samll value of u_{max}. However, the accuracy of the approximation resulting from the use of the second sampler with one half the base period, is within the instrumentation tolerance.

The signal out of G_1 and the derivative of the signal out of D_1 are included in both Figs. 6 and 7. The filtering which is apparent on these signals is basically responsible for the close check of the approximate solution (with $n=2$) and the exact solution.

The accuracy of these approximating results for the low n and u_{max} values in

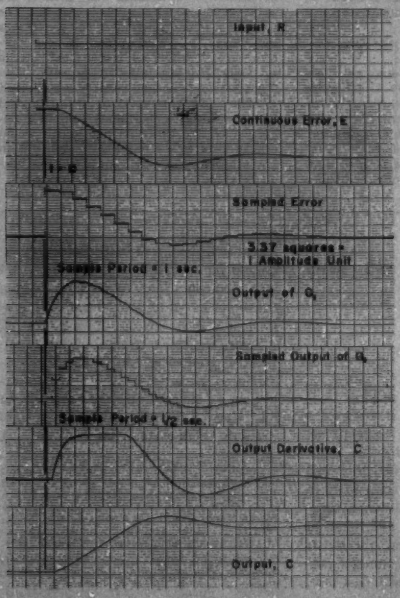


Fig. 6 (left). Approximate system response, $n = 2$, $u = \infty$, for illustration

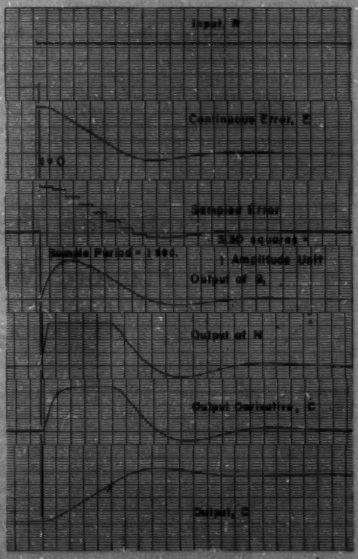


Fig. 7 (right). System response for illustration

this particular example must be generalized with care. A less severe nonlinearity, such as the saturation of a d-c generator, would tend to increase the accuracy of the approximation, while shorter time constants in the linear portion of the system would have the converse effect.

Modifications and Extensions

The application of equations 7 and 8 to a system such as is illustrated in Fig. 1, but with a nonstep input function, a multistorage nonlinearity, and a transport delay, follows the procedures discussed for the methods developed in a previous paper.[1]

Equation 9 is equivalent to equation 8 except for the evaluation of $E(t)$ at the high sampling rate. Therefore intersample values may be investigated by varying the value of n in equation 9.

A system containing time-varying parameters can be treated as suggested in reference 1, to the practical extent that the varying parameter influences just a few terms in a reduced equation, such as in equation 11.

Conclusions

Methods for obtaining the numerical solution of nonlinear sampled-data control systems, as illustated in Fig. 1, are described. The use of these methods makes it possible to obtain solutions to important nonlinear control system problems which cannot be solved by conventional methods or those described in reference 1. For finite n the solution is approximate. The methods are general in application and they permit any degree of approximation. As the accuracy of the approximation increases, the labor required to derive the solution increases, as expected.

The ease with which these methods are applied changes little with the relative complexity of the description of N and the form of $G_1(s)$. Various $D_1(s)$ functions, however, may increase the number of nonnegligible $\bar{d}_{n-p}$ terms in equation 8 thus varying the extent of one of the three summations. An actual problem evaluation requires only the use of standard z-transform tables and systematic calculations.

References

1. Analysis of Nonlinear Sampled-Data Control Systems, Part I, E. Kinnen, J. Tou. *AIEE Transactions*, see pp. 386–90 of this issue.

2. The Analysis of Nonlinear Sampled-Data Control Systems, E. Kinnen. *Ph.D. Thesis*. Purdue University, Lafayette, Ind., 1958.

3. Digital and Sampled-Data Control Systems (book), J. T. Tou. McGraw-Hill Book Company, Inc., New York, N. Y., section 6.7, 1959.

4. Analysis of Sampled-Data Feedback Control Systems with Finite Sampling Duration, J. Tou. *AIEE CP 58-801* (available on request).

Joint Discussion

Discussion of Papers 59-828 and 59-829

Thomas J. Higgins (University of Wisconsin, Madison, Wis.): The authors of these two papers have written a clear, easily followed account of an excellent numerical technique for analysis of nonlinear sampled-data systems. The numerical examples and experimental data confirm well the application and correctness of the general theory.

Keisuke Izawa (Tokyo Institute of Technology, Tokyo, Japan): The papers under discussion are an outstanding contribution to the theory of nonlinear sampled-data control systems. The proposed sequence method and the modified one with the introduction of multirate sampler are useful for tracing the transient responses of the control systems.

The G_1's in the papers might include the equalizer to achieve better control behavior. The design of the equalizer would be of particular interest to the practicing engineer. Although the introduction of the multirate sampler simplifies the analysis it might be of great help to discuss the adequate value of n of the multirate sampler.

E. Kinnen and J. Tou: The authors appreciate the comments made by Dr. Higgins and Dr. Izawa. The inclusion of an equalizer in G_1 to achieve better control behavior is certainly a valuable suggestion. In fact, work is now being undertaken to extend the techniques described in this paper to the design of equalizers, both continuous and discrete, for compensating nonlinear sampled-data systems. The results will be reported in a future paper.

The second comment made by Dr. Izawa brings up a good point. However, the choice of a suitable value of n for the fictitious sampler in Part II is dependent upon many factors, such as the properties of G_1, the sampling period T, and the nature of the nonlinearity. A detailed discussion of this point would require a whole paper. The main idea is to approximate the continuous input to the nonlinear element N by a staircase signal, which is more amenable to mathematical analysis. A general statement for the minimum value of n is difficult to make. Roughly speaking, if G_1 is a good low-pass filter, a fictituous sampling frequency of five times its bandwidth would usually lead to a satisfactory result. With the sampling period T specified, the adequate value of n can thus be estimated.

Cathodic Protection Applications at the Hanford Works

C. S. BUCHOLZ
MEMBER AIEE

THIS PAPER describes existing, presently planned, and experimental impressed-current cathodic protection (CP) installations at the Hanford Atomic Products Operation. The Hanford plants are located in the south-central part of the state of Washington. Climatic conditions are semiarid and the plant area is typical desert country. The soil is predominantly glacial deposits ranging from fine sand to gravel mixed with larger boulders. This water-washed material has a very high resistivity, from 5,000 to 100,000 ohm-centimeters, although there are local strata of hardpan and alkali of lower resistivity. In general, the soil is only mildly corrosive to ordinary materials. Most of the original cast-iron, galvanized-steel, and carbon-steel underground piping installed in 1943 and 1944 is in good condition. As discussed herein, the direct burial of unprotected stainless-steel pipe lines has not been successful.

Some serious corrosion of large galvanized grounding cables has been experienced because of the wartime expedient of using available materials, in this case a lightly galvanized, high-strength (high-carbon) steel strand. A

Paper 59-821, recommended by the AIEE Chemical Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing May 22, 1959.

C. S. Bucholz is with the Hanford Atomic Products Operation, General Electric Company, Richland, Wash.

Fig. 1. Steel grounding cable after burial in sandy soil for approximately 14 years. The galvanizing has been consumed in the dark areas

typical example is shown in Fig. 1. Extra-heavy galvanizing is specified for new construction. Failure of underground steel piping has been caused by near-by bare-copper ground conductors and, for this reason, copper is no longer used in the vicinity of other buried metals.

Process water for the Hanford facilities is pumped from the Columbia River, filtered, and treated. Resistivity varies with the season of the year and treatment, but has an average range of 3,000 ohm-centimeters at 40 F (degrees Fahrenheit) to 8,000 at 170 F. Sanitary water is taken from the process water supply (or pumped from wells near the river) and chlorinated. It is about 3,400 ohm-centimeters at 70 F. These waters are quite aggressive to carbon steel.

Exclusive of a few specialized applications, the four subjects discussed herein essentially cover the scope of CP work at Hanford. A general knowledge of CP principles on the part of the reader is assumed.

Cathodic Protection of Underground Stainless-Steel Process Lines

Buried stainless-steel pipelines are used extensively for transport of liquid process materials between separations plants, waste storage tanks, and other facilities. The lines are from 1 to 4 inches in nominal diameter and are mostly of the *18-8* grades of stainless steel. All joints are shielded-arc welded. Some of these process materials are highly radioactive and any leakage into the surrounding earth could have serious consequences. To prevent such incidents, lines installed since 1948 are encased in concrete ducts or steel pipe (except those carrying low-level radioactive wastes). The ducts are arranged to drain into catch tanks from which any leakage can be removed safely.

Approximately 5 miles of pipe for the original separations plants were installed (1943–44) directly in the earth with no protective covering or coating. Several leaks in these lines had occurred by 1947 but corrosion was not seriously considered as the cause at that time. During that same year a new line was laid as a part of a process change. The excavation was partially backfilled, and the pipe left exposed to air and moisture for about 2 months. No pressure could be maintained when testing was attempted because the new line had become literally honeycombed by corrosion. This surprising incident definitely established the fact that, under certain conditions, stainless steel does corrode rapidly in Hanford soils. Corrosion starts in small round pits 1/8 to 3/16 inch in diameter. The pits or craters, when active, contain a moist black corrosion product, and often apparently migrate down and around the pipe giving the metal a worm-eaten appearance.

After a thorough investigation of the problem, CP was recommended as the most practical method to prevent further deterioration of underground stainless lines, and late in 1947 protective systems were installed on the 5 miles of existing piping. About half of these lines are still in use, and only five leaks have been detected since the inception of the CP program. Three of these are thought to be, at least in part, caused by factors other than external corrosion, i.e., damage by heavy machinery, abnormal stress at a diversion box, and other damage to an unusually light-gage old line.

A total of approximately 15 miles of encased or bituminous-mastic-coated stainless-steel process lines has been constructed since 1948, all cathodically protected. Parallel lines are often placed in a common concrete duct and single runs are sometimes contained within a larger carbon-steel pipe. In either case, the stainless pipe is supported at intervals on insulating collars so that the metal is not in contact with any conducting media. Cathodic protection is used to obviate the possibility of corrosion in the event that process solutions or water should enter the encasements.

Anodes for these CP systems are made

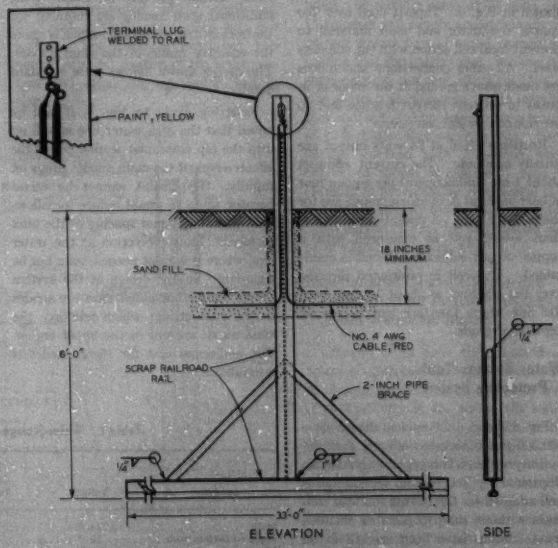


Fig. 2. Anodes used for protection of stainless-steel pipelines

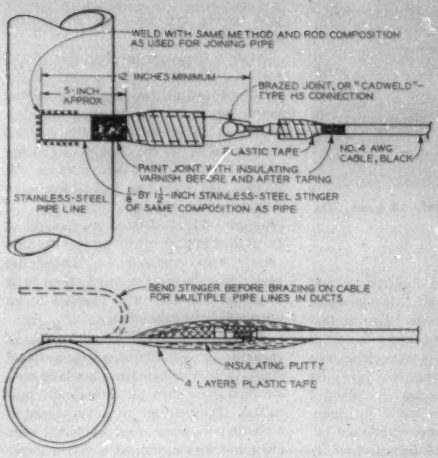

WELD WITH SAME METHOD AND ROD COMPOSITION AS USED FOR JOINING PIPE
12 INCHES MINIMUM
5-INCH APPROX
BRAZED JOINT, OR "CADWELD"-TYPE HS CONNECTION
PLASTIC TAPE
NO. 4 AWG CABLE, BLACK
PAINT JOINT WITH INSULATING VARNISH BEFORE AND AFTER TAPING
STAINLESS-STEEL PIPE LINE
1/2-BY 1/2-INCH STAINLESS-STEEL STINGER OF SAME COMPOSITION AS PIPE
BEND STINGER BEFORE BRAZING ON CABLE FOR MULTIPLE PIPE LINES IN DUCTS
INSULATING PUTTY
4 LAYERS PLASTIC TAPE

Fig. 3 (left). Cathode connection to stainless-steel lines

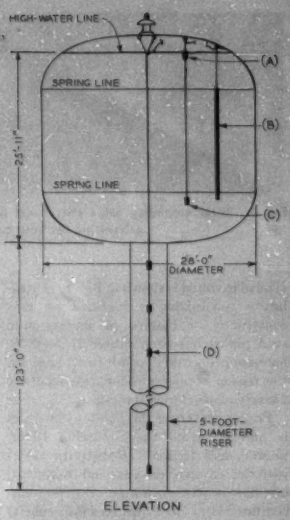


Fig. 4 (right). Anode configuration for an elevated water-storage tank

A—Six $1^1/_4$- by 12-inch top anodes
B—Six 1-inch by 15-foot main anodes on 9-foot radius
C—Six $1^1/_4$- by 12-inch stub anodes on 5-foot radius
D—21 $1^1/_4$-by 12-inch riser anodes, 5-foot 6-inch spacing

from railroad rail because scrap rail is readily available and a large effective area can be obtained. A typical anode detail is shown in Fig. 2. Anodes are spaced 150 feet apart alternately on both sides of the pipeline, and the same distance laterally from the line. Close spacing is desirable in this high-resistivity soil. The method of attaching the cathode lead to lines is shown in Fig. 3. Tape is used over the copper conductor and weld material to prevent local cell action with the stainless steel. All other connections and splices are made above ground at the anode or in boxes to prevent current leakage and to provide convenient test points.

Rectifiers rated at 24 volts output are usually adequate. The current averages about 1 ma (milliampere) per square foot of bare pipe but is, of course, much less for the encased lines. A potential of not less than −0.60 volt as measured with a copper–copper-sulphate half-cell is maintained. The cell is positioned directly over the pipeline, or about 10 feet from the line where lines are less than that depth from grade.

Water-Storage-Tank Protective System

Fig. 4 shows a CP system design for a 100,000-gallon elevated storage tank for sanitary water. This type of tank has an ellipsoidal top and bottom section and is well adapted for CP having rounded contours without sharp corners or shielded areas. Many have been erected in all parts of the United States including several others at Hanford. In the plant areas, these tanks are used for fire protection or other emergency water supply and are therefore kept completely filled.

The conventional anode layout for tanks of this kind employ a number of main anodes in the bowl and sectionalized anodes extending the full length of the riser. Except possibly for small tanks, additional stub anodes are required for the bottom area of the bowl which is inside the effective radius of the main anodes. The design shown follows these standard details.

Referring again to Fig. 4, it will be noted that the high water line is well up into the top ellipsoidal section where the effectiveness of the main anodes drops off rapidly. The mains cannot be raised because to do so would reduce the allowable 5-foot minimum spacing to the tank surfaces. Thus protection at the water line, where it is most needed, may not be adequate. In the design of the system shown, the author added short top anodes to the stub strings which reinforce the main anode current in the upper region. The additional cost for these anodes is insignificant.

The interior of the tank was painted just prior to installation of CP in November 1958. Six 1-inch-diameter by 15-foot-long main anodes, six top and six bottom $1^1/_4$- by 12-inch stub anodes, and 21 $1^1/_4$- by 12-inch riser anodes, all of high-silicon cast iron, were used. A dual CP rectifier was installed with one section for the main and the other for the riser and stub anodes. With the addition of a rheostat in the stub feed, the voltage on each of the three anode assemblies can be individually controlled. Design anode resistance, and voltage and current requirements were close to actual operating values. Data taken after the system had become stabilized are shown in Table I. Final testing and adjustments have not been completed to date because of inclement weather.

Protecting a Reactor-Containment Vessel

A common design feature of reactors to be located in populated areas is the large

Table I. Water-Storage-Tank Protective System

Anodes	Voltage	Total Current, Amperes	Actual Resistance, Ohms	Calculated Resistance, Ohms
Main	14.6	11.6	1.26	1.08
Top and bottom stub	10.0	1.1	9.10	9.50
Riser	14.9	7.1	2.10	1.25

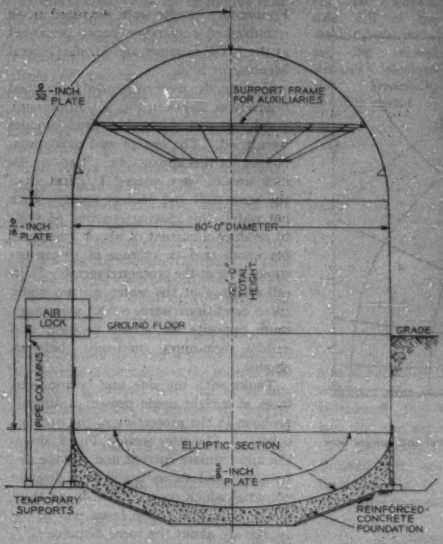


Fig. 5 (left). Section through welded-steel reactor-containment vessel

Fig. 7 (above). Experimental installation in a steel retention basin showing the wood brackets and graphite anodes

steel sphere (or similar form) entirely surrounding the reactor. The objective is, of course, to hold within the shell any radioactive materials that may result from a reactor mishap. Because they are pressure vessels, the underground portion should be protected from corrosion.

A containment vessel for a research reactor now under construction at Hanford is shown in Fig. 5. The design does not provide for any corrosion allowance on the plate thickness. Nearly half of the structure is below grade, and the underground surfaces were covered with an asphalt-impregnated fiber membrane, 5/32 inch in thickness, applied with hot asphalt. The concrete foundation was poured after fabrication of the lower portion and after the membrane was in place.

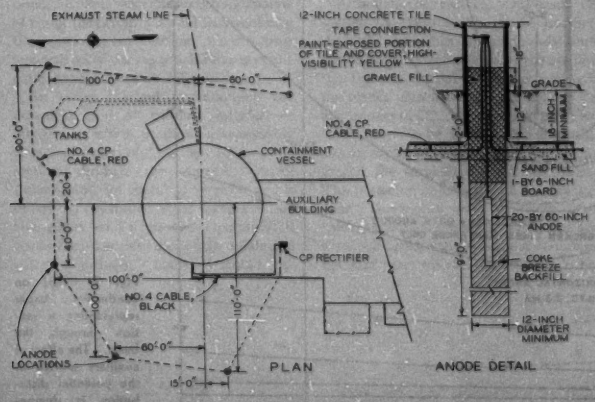


Fig. 6. Layout and anode detail for protection of the containment vessel shown in Fig. 5

Cathodic protection will be applied immediately after backfilling. Current will be supplied from six 2- by 60-inch high-silicon iron anodes with coke breeze backfill, the first installation of this kind at Hanford. Graphite anodes would also be suitable for this application. The anode layout is shown in Fig. 6. Irregular anode spacing was necessary to avoid underground service lines and to provide some protection to this piping and to auxiliary tanks. A rectifier rated at 48 volts and 15 amperes will supply approximately 10 ma per square foot of metal surface assuming that the coating may, in time, become only 90 per cent effective. Fig. 6 also shows a combination anode installation and test station made from a length of 12-inch-diameter concrete tile. Sand backfill covered with creosoted wood boards, as shown, is commonly used at Hanford to provide some mechanical protection for directly buried electric cables.

No operating data for this system will be available for several months.

Experimental Cathodic Protection Installation for Reactor-Cooling-Water Retention Basins

Effluent water from the Hanford reactors is discharged into retention basins so that, in the event of an abnormal level of radioactivity in the water, the radioactivity can be allowed to decay before the water is returned to the Columbia River. Several very large steel tanks are used for this purpose. The water is thermally hot and ordinary paint coatings are not suitable for use on the inside surfaces of these basins. Two open tanks, constructed in 1952 for use with the *105-C* re-

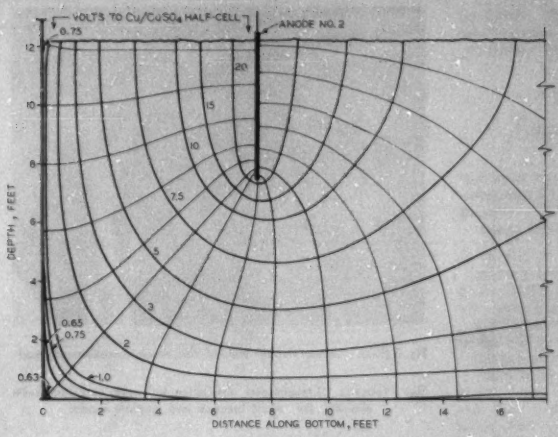


Fig. 8. Cross section of the basin at an anode location showing equipotential and current flow lines

Water temperature = hot
Rectifier voltage = 35
Anode current = 1.99 amperes

actor, are 330 feet in diameter, have straight 16-foot sides, and a capacity of 9 million gallons each. The original paint had begun to fail after about one year of operation.

It was then decided to investigate the possible application of CP to these basins and, in the summer of 1954, an experimental system was installed in one of the tanks to operate on a small segment of the side wall. This work was continued by the author through July of 1957. Preliminary results were discussed in an unpublished conference paper presented at the AIEE Summer and Pacific General Meeting in 1956.

Four anodes were suspended from wood brackets as shown in Fig. 7. Aluminum, graphite, and high-silicon-iron anodes were tried. The last proved most satisfactory for this application. Two- by 60-inch anodes were spaced 7½ feet from the tank wall and 15 feet apart. With hot water, this spacing required 35 volts to produce a current of about 2 amperes per anode and an average of 10 ma per square foot at the protected metal. Half-cell readings at the water surface under these conditions were −0.82 volt maximum opposite an anode position and −0.70 minimum mid-way between anodes.

Tanks with the side and bottom surfaces at a right angle present a difficult problem in the protection of the extreme corner. The cross section, Fig. 8, shows the approximate current and voltage distribution and shows the serious drop in potential which occurs at the lower third of the tank wall. The lower end of the anodes are about the same distance (7½ feet) from the tank bottom as the radial spacing from the wall. Little would be gained by using longer anodes because a large part of the current would flow to the bottom in the region directly beneath them.

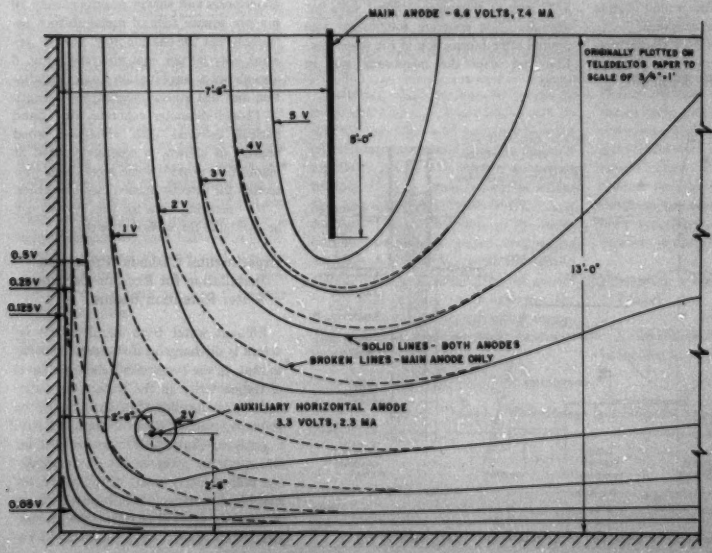


Fig. 9. Plot on conductive paper representing a section through the basin. The effect of auxiliary anodes on the potential distribution is demonstrated

It was realized from the beginning that the corner areas were probably not receiving adequate current for complete protection, but it was hoped that the effects of prolonged polarization and possibly some mineral deposition would eventually force the current to progressively lower surfaces. This did not materialize. There was no significant change in the potential distribution after nearly 2 years of operation.

Further effort was then made to investigate the degree of coroner protection by means of Corrosometer probes and the use of a one-foot-square test plate. This only confirmed what was already known, that the lower third of the wall was not receiving sufficient current. The Corrosometer showed complete protection at the top and half-depth probe locations but the rate of corrosion at the bottom was not much improved over that at an unprotected location. Currents to the test plate were about 6 ma per square foot at the bottom, 12 at half-depth, and 20 at the surface.

The conventional design methods for bowl-shaped tanks containing quiet water of appreciable mineral content cannot be applied to these basins. None of the essential factors are present, i.e., the tanks have a sharp, square side-bottom intersection, much more current is required to maintain the polarizing film of hydrogen with rapidly flowing water, and there is little, if any, mineral deposition at the current density required for protection. It is quite possible that adequate corner protection could have been attained had any one of these factors been favorable.

Fig. 8 is a plot of the actual measured half-cell potentials taken at various locations. Curves were drawn through points of equal potential. To aid in visualizing the pattern of current flow, one can draw a second set of curves which interesect the equipotential lines at right angles.

An easier method for mapping the potential field at corners, interference areas, etc., which produces a reasonably accurate "picture," is as follows. A section through the current source (anode) and the metal to be protected is drawn on resistance paper using conductive paint, and a small d-c potential is applied across this analog of the anode and cathode surfaces. Lines of equal potential can then be traced on the paper with a probe connected to a sensitive voltmeter. Fig. 9 is a plot made in this way for the same cross section. Note that the shape of the potential curves (broken lines) are almost identical with those of Fig. 8, although the numerical values now have no particular significance.

Fig. 10. Interior side wall of the drained basin showing the mineral deposits obtained with abnormally high current

The solid curves in Fig. 9 illustrate the effect of introducing a horizontal auxiliary anode in the corner area. Many combinations of position and fraction of main anode voltage applied to the auxiliary anode are possible. The arrangement shown is the most successful of several made by the author. Note how the equipotential lines are straightened down into the coroner, which is a great improvement over the distribution with the main anode only (broken lines). Nevertheless, the extreme corner inevitably remains at a reduced potential.

Although the cost of a full-scale CP system for these tanks can be justified, the lower few inches of the side walls could not be fully protected without the application of some type of insulating coating in the corner area which largely defeats the advantages of CP. The original paint has since ceased to be of any value and other studies are being made in an effort to prolonged the serviceability of the basins. The corrosion rate is about 10 mils per year. One-eighth inch of extra plate thickness was provided in the design and thus the tanks should have a minimum life of 12 years with no protection. It has been shown that, for new basins, the cost of providing the necessary extra metal thickness for any desired life is more economical than periodic painting or the installation of CP.

The author had been curious during the course of the experimental work as to why no mineral deposits could be observed on the tank wall. At the conclusion of the work it was decided to move two of the anodes to within about 2 feet of the wall in an attempt to force a precipitate. The current from each anode at this location was increased to 3.2 amperes. The photograph, Fig. 10, was taken about 3 weeks later when the tank was drained. A heavy localized deposit of calcium carbonate had formed on the tank wall directly behind the anodes which appears on Fig. 10 as the two white rectangular areas. However, calcareous deposition occurs only at a current density of about ten times that necessary to stop corrosion.

The other irregular light-colored areas in Fig. 10 are patches of the original paint. Presumably the paint film, somewhat thicker in these areas, had not deteriorated before CP was applied, and CP had halted further breakdown of the paint. The dark surfaces in the illustration are, of course, corroded metal and all surfaces outside the influence of the protective current are uniformly of this appearance. The contrast is not as strikingly apparent as formerly because the CP installation was inoperative for several months before the photograph was taken.

A Solid Rotor A-C Generator for High-Temperature Electrical Systems

J. T. BATEMAN
NONMEMBER AIEE

RELIABILITY has always been the primary requirement for aircraft generating systems. The rapid development of high-performance aircraft has continually demanded greater reliability along with increased operating temperatures. These two conditions are not directly compatible. As temperatures increase the choice of insulations and lubricants is narrowed to a few which, in general, have inherently lower mechanical properties than the conventional lower temperature materials. During high-temperature operation these materials, as well as the electromagnetic materials, suffer additional loss in physical and electrical properties. The desired increase in reliability must be achieved despite these obstacles introduced by the increased temperatures.

A development program aimed at generating equipment for 600 F (degrees Fahrenheit) ambient temperature has been in progress since March 1958. At the outset of this program it was recognized that the most painstaking development and application of high-temperature materials might still result in a generator which fell short of acceptable aircraft reliability. For this reason, a vital part of this development program has been the investigation of unusual types of generators where there are fewer components and where the stress of critical components is reduced.

This paper presents a brief survey of high-reliability generator configurations. The generator selected for the 600 F development program is described and its characteristics are related to conventional equipment. Bearing and insulation work on the 600 F program is described in other related papers.[1,2]

Objectives

The 40-kva 120/208-volt 0.75-power-factor 3-phase 400-cycle 8,000-rpm generator rating that has been specified in this development program is typical of the majority of current military and commercial a-c aircraft generators. Environmental specifications reflect the more severe conditions which would be encountered in a Mach-3 aircraft.

Ambient temperature	600 F
Vibration	20 g
Shock	50 g

Out of 500 hours of life, 20 hours are specified at the ultimate condition with the remainder at more conventional temperatures.

To maximize reliability under these severe environmental conditions, several additional design objectives were set.

1. Use a minimum number of parts, especially where high electrical or mechanical stresses will occur.
2. Remove all windings, rectifiers, and other electrical components from the rotor. This will also eliminate brushes, slip rings, and commutators.
3. Utilize configurations where the mechanical strength of the electromagnetic iron parts will add to the structure and ruggedness of the generator.
4. Select electromagnetic circuits which would allow rectifiers, capacitors, and other sensitive components to be relocated in a more favorable environment if necessary.

These approaches for achieving the greatest possible reliability were all tempered by the fact that lightweight design is an extremely important requirement in aircraft generating systems. The known electromagnetic configurations which would meet the high-reliability design objectives appeared to be somewhat heavier than conventional salient pole wound rotor designs. This was not considered an insurmountable problem since these high reliability configurations did not have the benefit of intensive creative design study directed toward their weight and performance optimization. With the expectation that significant improvements could be made, the investigation of new generator configurations was directed toward those having the greatest inherent reliability.

Permanent Magnet Alternator

The removal of all windings from the rotor could be accomplished with a permanent magnet type of generator with saturating windings in the stator providing voltage control. Because the excitation system serves to depress the output voltage, this type of generator would deliver overvoltages in the event of a failure of the regulating system. If an internal fault occurred in the generator, no method short of mechanically disconnecting the generator would be certain to prevent a fire. Only magnet materials having a high energy product would be considered for lightweight aircraft design. Above 600 F these magnet materials are gradually demagnetized.

Inductor Alternator

The inductor alternator has already distinguished itself as a generator configuration which meets all of the design objectives for high reliability. A brushless d-c generator has been built using the basic inductor electromagnetics with only a rotating iron configuration on its rotor. The 800-cps power generated by these electromagnetics is rectified to provide 100 amperes, 30 volts direct current at the terminals of the generator. This brushless d-c generator is now in use on a modern, high-performance interceptor aircraft.[3]

The basic principle of the inductor alternator in all of its varied configurations is well described in the literature.[4-6] In spite of this, it should be helpful to describe briefly, the double-ended, homopolar inductor configuration which is the basic design around which much of this study centered. This can be done conveniently with reference to the conventional salient pole synchronous generator pictured in Fig. 1(A). The alternately north and south magnetic poles of this rotor are induced by d-c field windings placed around each pole piece. Since these windings encircle each pole, they must rotate with the rotor and, therefore, field current from a stationary source must be supplied through some sort of sliding contacts or brushes to the rotor. This conventional generator can be converted into a homopolar inductor machine by the changes pictured in Fig. 1(B). The rotor body and the stator are lengthened, but the pole pieces themselves are unchanged. The north poles are all placed at one end and the south poles are All placed at the other end, but their angular relationship is unchanged. The modified rotor would still work with the original field coils, but this is unnecessary.

Paper 59-848, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 23, 1959; made available for printing April 30, 1959.

J. T. BATEMAN is with General Electric Company, Erie, Pa.

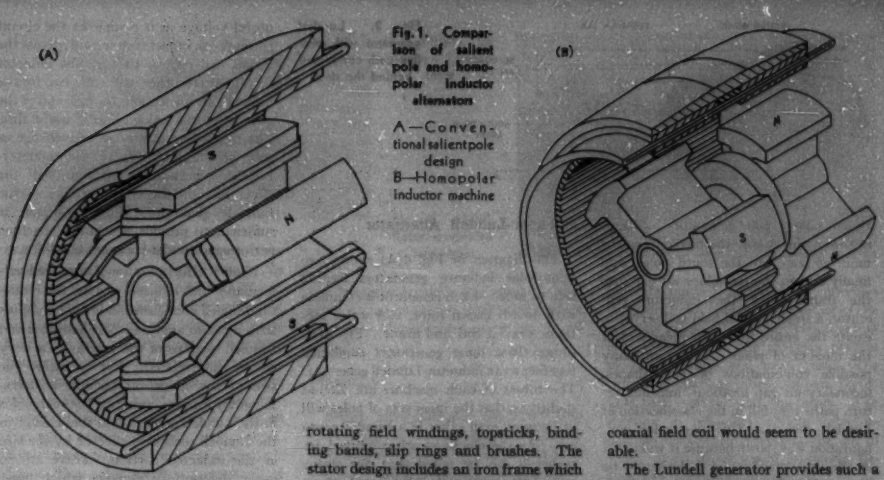


Fig. 1. Comparison of salient pole and homopolar inductor alternators

A—Conventional salient-pole design
B—Homopolar inductor machine

All of the magnetic flux which passes from the north poles to the south poles internal to the rotor must go through the center portion of the shaft. A single, coaxial field coil around the shaft at this point will link all of the flux and supply all of the necessary excitation. This coil is not required to rotate and it can therefore be fixed in the stator eliminating the need for brushes or other sliding contacts.

It is evident from Fig. 1(B) that the flux must travel axially both in the center of the rotor and outside the stator core. In order to accommodate this flux, solid steel paths are provided at both points. The coaxial coil is required to supply only as many ampere-turns as originally supplied by two of the separate coils of Fig. 1(A). In the case of a 6-pole generator this represents a significant reduction in excitation power. The sketches of Fig. 1 might suggest that the inductor machine is more than twice as large as the conventional machine. This is not the case. The inductor is usually heavier, but the typical increase is only about 60%. The lower excitation power requirements and elimination of current collection provisions will actually result in a reduction of system weight or volume for some applications.

The actual rotor used in the 100-ampere brushless d-c generator is shown in Fig. 2, along with the stator. The coaxial field coil can be seen inside the stator. This configuration satisfies all of the design objectives for greater reliability. The rotor has been reduced to a minimum number of parts by the elimination of rotating field windings, topsticks, binding bands, slip rings and brushes. The stator design includes an iron frame which not only carries flux but also adds to the frame structure of the generator. Because all of the electrical circuits are located on the stator, remote location of associated rectifiers and capacitors is possible. Oil cooling may be accomplished with a simple and reliable circuit of oil tubing across the outside of the frame. For 40-kva 6-pole generators, however, the additional weight involved in the usual double-ended homopolar inductor design would indicate that improved variations of the basic configuration should be considered.

Lundell Alternator

The weight disadvantage in the homopolar inductor generator is largely explained by the fact that the north and south poles are at opposite ends of the rotor and no part of the stator sees a full north to south variation of flux. With this half-wave flux, the stator iron is used to only one half of its capability and accordingly the machine is heavy. A rotor configuration which would present alternate north and south poles to a single stator while retaining the feature of a coaxial field coil would seem to be desirable.

The Lundell generator provides such a configuration. North and south polarities are maintained at opposite ends of the rotor through the action of a coaxial field coil. The north and south poles are not positioned at opposite ends of the rotor, however. North and south poles are extended from each end toward the center of the rotor so that north poles project between south poles and vice versa. This presents an array of alternate north and south poles to the stator and provides for a full-wave flux variation and optimum stator electromagnetic weight.

One of the many possible arrangements of Lundell electromagnetics is shown in Fig. 3. One of the marks of a brushless Lundell configuration may be seen in this sketch. Two secondary or parasitic air gaps are always required to complete the rotor flux circuit and yet allow the use of a stationary coil. In the configuration shown, flux travels along the frame and crosses one secondary air gap into the north pole end of the rotor. The flux then travels out along the cantilevered north pole fingers, emerging from the north pole faces and crossing the main air gap into the stator. The flux follows the same paths in the stator as it does in conven-

Fig. 2. 30-volt 100-ampere d-c generator disassembled rotor and stator showing inductor construction with coaxial field coil visible in stator

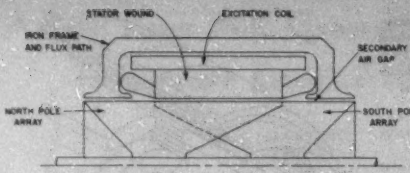


Fig. 3. Lundell generator with rotor flux path completed behind the stator

tional salient pole generators. Emerging from the stator, the flux crosses the main air gap into the south poles, and finally crosses the secondary air gap into the frame. Like the inductor generator, a single coaxial field coil serves to excite the entire flux path regardless of the number of poles. There are many possible combinations of pole shapes, secondary air gap locations, and flux return paths that fall in the classification of Lundell designs. This particular configuration was shown because it will assist in the understanding of a design to be discussed later.

The Lundell generator has one inherent characteristic which affects its weight as compared with conventional generators. The rotor of a Lundell generator tends to be a very closely intertwined 3-dimensional configuration where the leakage between north and south poles is about four times the rotor leakage found in conventional salient pole wound rotor generators. Part of this is due to the absence of windings on the Lundell poles so that the full mmf (magnetomotive force) exists between all north pole and south pole surfaces. If a short rotor flux circuit is designed, then the rotor leakage will be high because of the proximity of rotor parts. With this high leakage, the cross section of the rotor iron must be increased to accommodate the leakage flux and the rotor weight will be accordingly increased. If a long rotor flux circuit is designed in an effort to separate the parts and reduce the leakage, the rotor weight will still be high because of the length of the flux path. Brushless Lundell machines will generally be 30% heavier in electromagnetic weight.

The high rotor leakage also has an adverse effect on generator transient performance. Rotor leakage flux represents stored energy which will tend to cause voltage overshoots upon load removal. Rotor leakage reactance also increases the open circuit time constant of the generator. Lundell configurations clearly satisfy the design objectives for high reliability but still indicate a need for further improvement in the direction of lighter weight and improved transient performance.

Inductor-Lundell Alternator

The diagram in Fig. 4(A) shows two homopolar inductor generators located side by side. Each generator is complete with double ended rotor, two stator sections, coaxial coil and frame. Fig. 4(B) shows these same generators combined together as an inductor-Lundell generator. The rotors of each machine are altered slightly so that the inner sets of poles will fit between each other in a Lundell manner. As the machines are combined together it will be noted that one stator section is completely eliminated along with a significant portion of the stator conductors and conductor end turns. The inductor generators when combined together in this inductor-Lundell configuration are significantly lighter than an inductor generator alone.

The inductor-Lundell configuration bears a certain resemblance to the Lundell design sketched in Fig. 3. The two parasitic air gaps have been replaced by working gaps with inductor power being generated at each end of the machine. The Lundell pole-to-pole leakage still exists and is still greater than that generated in conventional machines. However, the Lundell pole configurations naturally extend in opposite directions to minimize leakage without necessarily increasing the length and weight of rotor iron. More significantly, the effect of Lundell leakage on over-all generator characteristics is drastically reduced. The flux which leaks between Lundell poles fails to generate voltage in the center stator section and so affects generator characteristics as seen by that portion of the stator. However, this same leakage flux generates useful voltage as it completes the circuit through one inductor generator, across the frame, and returns through the inductor section at the other end. As seen by the inductor stator sections on each end of the generator, this leakage flux is useful flux. The total effect of the Lundell leakage as seen at the terminals of the inductor-Lundell generator is less than half the effect seen on the characteristics of the Lundell generator. This reduction is sufficient to provide acceptable machine performance under transient conditions.

While it offers promise of improved performance and reduced weight, the inductor-Lundell configuration is very complex to design and analyze. The inductor and Lundell sections of the machine are very closely interrelated. For example, the size of the induction portion is determined by leakage occurring between poles in the Lundell sections and the leakage in the Lundell section is affected by the size of the inductor portion. There are so many different parameters which must be solved by trial and error that effective and accurate designs cannot be obtained in a reasonable time by hand calculation. The inductor-Lundell design formulas have been programmed for use on an International Business Machines Corporation *650* computer. The program incorporates many loops which allow the computer to balance interrelated parameters and to investigate possible variations leading to the lightest design within the limits of the input requirements. The design work on the computer has shown that inductor-Lundell electromagnetic weights are generally less than the weights of either the basic homopolar inductor or Lundell configurations for conventional speeds, frequencies, and ratings. The inductor-Lundell configuration, like the basic inductor and Lundell machines, may be lighter than conventional salient pole designs for many of the unusual speeds, ratings, and duty cycles that are specified from time to time in aircraft applications.

An inductor-Lundell rotor is shown in Fig. 5. This rotor is from a generator which was built to the specifications of the

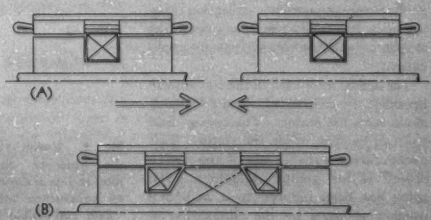


Fig. 4. Comparison of inductor and inductor-Lundell generators

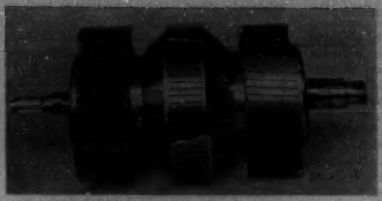


Fig. 5 (left). Inductor - Lundell rotor

Fig. 6 (right). Rotor assembled into stator

600 F development program except that high-temperature component materials were not used. The center Lundell sections of this rotor are solid iron while the inductor sections at each end are assembled from iron laminations. Other configurations are possible where laminations are used throughout. The center section of the shaft must be nonmagnetic material or flux will leak directly between Lundell poles through the shaft. Amortisseur bars are included at the pole faces. The rotor as shown meets all of the design objectives for high reliability. Fig. 6 shows the inductor-Lundell rotor assembled into the oil-cooled stator.

Results of both analytical work and testing of inductor-Lundell machines have shown that electrical characteristics are comparable with conventional generators. Generator characteristics, voltage balance under unbalanced loads, waveform, efficiency, and short-circuit capabilities have all been better than specifications. In several respects, the dual nature of the inductor-Lundell generator is providing new methods of design improvement which should challenge the capabilities of conventional aircraft generator designs. The increase in reliability of the inductor-Lundell configuration suggests its use in many applications where extremely dependable operation is required.

Summary

Severe environmental conditions and the continuing need for greater reliability have suggested the further development of unconventional generator configurations which have fewer components and where stress on critical components is reduced. Inductor and Lundell generators satisfy this design objective since they involve simple iron configurations on the rotor with no rotating electrical components. The homopolar inductor design in its basic form is 60% heaver in its electromagnetic weight at the most significant ratings and speeds. The basic Lundell design involves less weight penalty but has a high leakage flux in the rotor which adversely affects transient performance.

A configuration combining inductor and Lundell machines has been developed to overcome the disadvantages that these configurations exhibit separately. The inductor-Lundell combination provides generally lighter weights than either of the separate types. The Lundell leakage flux is put to work in the inductor portion of the machine and as a result, the generator characteristics compare favorably with conventional generators. The increase in reliability of the inductor-Lundell configuration suggests its use in many applications where extremely dependable operation is required.

References

1. Bearing Systems for High-Temperature Aircraft Generators, G. D. Bradley. *AIEE CP 59-865* (available on request).

2. Generator Insulator Systems Development for Hypersonic Aircraft, W. B. Pegg, R. L. Balke, F. M. Precopio. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, Sept. 1959, pp. 255–59.

3. A Brushless D-C Generator for Aircraft Use, J. T. Duane. *Ibid.*, vol. 77, Nov. 1958, pp. 367–72.

4. The Theory of the Inductor Alternator, J. H. Walker. *Journal*, Institution of Electrical Engineers, London, England, vol. 89, pt. 3, 1949, pp. 227–41.

5. Short Circuit Characteristics and Load Performance of Inductor Alternators, A. Maadi. *Ibid.*, vol. 94, 1947, pp. 102–17.

6. High-Frequency Alternators, J. H. Walker. *Ibid.*, vol. 93, pt. 2, 1947, pp. 67–80.

Discussion

Richard T. Smith (University of Texas, Austin, Tex.): The author is to be congratulated for proposing a novel solution to the problem of providing reliable generation at high temperatures with available materials. The new generator appears to have certain components of leakage which would be higher than for conventional designs. In particular, the peripheral end leakage may be quite significant, whereas, for many standard machines, this component is negligible. In addition, the differential and belt leakages in the Lundell Section will be higher than for the remaining sections. The author does not indicate that any special provisions are made for quadrature axis damping. It is clear that such damping can be achieved, however, perhaps more easily than for the conventional salient-pole construction.

K. M. Chirgwin (Jack & Heintz, Cleveland, Ohio): The author states that the electromagnetic weight of the homopolar inductor alternator is 60% greater than that of a conventional wound-rotor salient-pole alternator. This seems to be a very optimistic figure. Would the author indicate whether all or part of the weight of the steel yoke was excluded in making this comparison? Further details of the applications in which the homopolar inductor alternator results in a reduction of system weight and volume would add to the interest of the paper.

When comparing the excitation requirements of the homopolar and conventional alternators, the author concludes correctly that the homopolar machine needs less excitation power, but the wording could lead to a misunderstanding of the reason for this. If both types of machines had circular exciting coils, then the power requirements would be the same; because, although a 6-pole homopolar machine needs only one third of the total ampere turns, three times as much flux must be carried by the core on which the coil is wound, and consequently the mean length of one turn of the coil is three times as great. The reason for the lower excitation power requirements of the homopolar machine is to be found solely in the more favorable geometry of the exciting coil.

Turning now to brushless Lundell-type machines, the author states that these machines are generally 30% heavier in electromagnetic weight. If this refers to the type of machine[1] shown in Fig. 3, where the flux returns through the endbells and stator housing then this figure seems to be much too low, unless the weight of endbells and housing is omitted. If it is intended to refer to those brushless Lundell machines where the flux returns through the shaft[2,3] then the figure of 30% is too high.

In the summary the author seems to imply that the electromagnetic weight of his inductor-Lundell alternator is less than 30% greater than that of a conventional machine. Would the author quote the actual figure, or otherwise clarify this point?

In defense of the brushless Lundell machine it should be pointed out that the so called secondary air gaps are effectively

in series with the main air gap to both the exciting and the armature reaction ampere turns. So these secondary air gaps are only a penalty on the design to the extent that they must for mechanical reasons be made larger than is required for electrical reasons.

The author is correct in stating that the Lundell machine has a high rotor leakage flux, but he is incorrect when he states that this leads to increased transient reactance and increased voltage overshoot on removal of load. The voltage overshoot and transient reactance depend much more on the ratio of armature reaction ampere turns to exciting ampere turns than they do on rotor leakage flux. By suitable choice of magnetic and electric loadings of the stator it is possible to attain any reasonable value of machine transient reactance. The secondary air gaps previously mentioned tend to give the machine a "naturally" low transient reactance.

REFERENCES

1. A New System of Autonomous Power Supply for Railway Passenger Cars (in Russian), G. I. Shturman, E. A. Yakubaitis, A. P. Krogeris, V. V. Apsit. *Elektrichestvo*, Moscow, U.S.S.R., 1957, no. 3, pp. 39–43.

2. A Rugged Power Team. *Machine Design*. Cleveland, Ohio, May 29, 1958, p. 43.

3. Dynamo Electric Machines, A. Bekey, H. M. Robinson. *U. S. Patent no. 2,796,542*, June 18, 1957.

R. E. Smith, W. J. Shilling (Westinghouse Electric Corporation, Lima, Ohio): Certainly with the ever-changing requirements of modern aircraft electrical systems, the generator design engineer must continuously evaluate the advantages and disadvantages of every generator type known, with their possible variations, and any new type which he might conceive. It is very important, however, that the major advantages of the conventional salient-pole construction be carefully considered before choosing a less conventional type.

It can be agreed that the inductor and inductor-Lundell generators weigh considerably more than salient-pole generators. It can also be agreed that the performance is comparable to salient-pole generators, but certainly not equivalent for the normal range of speed and frequency. Basically, the inductor-Lundell generator must have at least 50% more stator iron weight, stack length, and leakage reactance than a conventional salient-pole construction and the stator copper loss and stator copper weight will tend to be approximately 50% greater. The increased losses result in greater system heat generation and lower efficiency. Besides the generator weight penalty, another weight penalty also results because of additional fuel and cooling required. The increased leakage reactance increases voltage overshoots and voltage unbalance. The 50% increase in the factors previously mentioned are basic. There are, however, additional performance problems for the inductor-Lundell generators which increase these factors considerably beyond 50%.

The author indicates that the inductor-Lundell design is *less than* 30% heavier than the conventional salient-pole machine. It would appear that the inductor-Lundell design would have about half of the penalty of the straight inductor type since the inductor-Lundell stack must be increased by 50% over the salient-pole machine versus 100% increase for the inductor machine. Also, the inductor-Lundell machine described requires two stack separations. This means one more added stack separation than for the inductor generator. Thus, it would appear that since the inductor machine has about a 60% weight penalty (per the paper), the inductor-Lundell machine would generally be *more than* 30% heavier in electromagnetic weight than a salient-pole generator for normal speeds and frequencies.

Although field excitation tends to be less in the solid rotor designs, it is not necessarily true that it always will be. In the inductor-Lundell design described, twice as much stator mmf is required than in the inductor design. Thus, the ampere-turn requirement of the stator coils of the inductor-Lundell machine is the same as that for four poles of a salient-pole machine (4 coils of Fig. 1). Depending on the particular design considered, the mean-length turn of the field coil of the solid rotor machine may be greater than the length of turn in an equivalent salient-pole machine. In general, it is expected that the field excitation of a 4-pole inductor-Lundell generator will be about the same as the excitation of a 4-pole salient-pole generator for the same field copper weight. Any difference will depend upon the difference of coil mean-length turn and coil temperatures.

Another item that should be considered in the inductor-Lundell design is leakage reactance or Potier reactance. In comparing this machine with the salient-pole design, it is seen that the total stator stack is at least 50% greater in the inductor-Lundell design and hence, has inherently more leakage reactance. To evaluate this leakage reactance, assume all the poles of the inductor-Lundell design are under the middle and one end-stack only. This leaves the other end-stack with no poles under it. The reactance in the two stacks with the poles under them would be the same as that in an equivalent salient-pole machine.

In the inductor-Lundell machine more reactance is introduced because: 1. slot reactance in the stack with no poles under it; 2. leakage reactance in the stator winding at the stack separations; and 3. leakage flux in the air where the poles were removed. The last two items may be calculated by assuming the length of the straight portion of the a-c coil end extensions are increased an appropriate amount.[1] (This may be determined after considering how the reactance is determined for end extensions.) Calculations of the leakage reactance were made on a sample inductor-Lundell machine. Load tests were later run which verified these calculations so far as they can be verified by load tests. These results indicate that the designer should expect about a 160% increase in leakage reactance (2.6 times reactance of an equivalent salient-pole machine) for a well designed a-c stack. To overcome this reactance increase at overload conditions, a weight penalty is required. This leakage reactance increase will increase the problem of voltage overshoot upon load removal. The reactance increase will also adversely effect machine constants such as X_d, X_d', X_q', X_d'', X_q'', X_0, and X_2. The effect of the last two constants is to increase the voltage unbalance problems.

As the author states, the "leakage flux" in the middle stack becomes useful flux in the two end portions of the machine. This flux then is only half as effective in developing voltage as the "useful flux" is. Thus, half of this leakage flux may be considered true leakage flux. In the salient-pole design the leakage flux is carried only a short distance from one pole to the next. In the inductor-Lundell design the leakage flux must be carried the length of the rotor and back again by way of the stator. This extra length of leakage flux path means extra steel with corresponding weight penalty. The end stacks must be increased to carry leakage flux which again penalizes leakage reactance.

In regard to voltage unbalance, it appears from Fig. 5 that it would be difficult to obtain quadrature axis damping. Meeting the voltage unbalance requirements of *MIL-G-6099A* should require a considerable sacrifice in weight and size not only because of the problem of obtaining quadrature axis damping but also because of the very high leakage reactance of this type of design. Laminations might be used on the rotor to simplify application of damper bars and to reduce pole face losses. However, this would destroy the advantage of ruggedness in the solid rotor design.

With respect to reliability, certainly the elimination of brushes, brush rigging, slip rings, and commutators increases the reliability of a generator. It was for this reason that a great deal of effort has been concentrated on brushless generators with integral a-c exciters and rotating rectifiers. Removing rectifiers and windings from the rotor does not necessarily contribute to increased reliability. Our reliability information indicates that the rotating field winding is perhaps the most reliable winding in generators. One reason for this is the fact that there is only a very small voltage (1 or 2 volts) present between adjacent turns since they are wound in series on a given pole and then all poles are connected in series. Further, the normal voltage to ground is typically 30 volts. In the a-c stator, however, the voltage of the main a-c winding to ground is 170 volts peak and the voltage between turns can be 295 volts peak. Applying field windings adjacent to these high-voltage stator windings might well reduce the reliability. The fact that the rotating field windings are under compression due to centrifugal force makes them much less susceptible to damaging vibration effects than a field on the stator. Rotating rectifiers have had exceptionally high reliability. In one application which has accumulated an appreciable amount of operating time, the mean-time to failure of the rectifiers is 60,000 hours based on 240,000 rectifier flight hours.

We would like to ask the author the following questions:

1. Assume, for example, a 40-kva 8,000-rpm 400-cps machine built to meet *MIL-G-6099*.
   (a). How would the inductor-Lundell generator weight compare with a

conventional 40-kva brushless salient-pole air-cooled machine weighing 70 pounds with integral exciter for class *C* operating conditions?

(b). What would the excitation requirements of the inductor-Lundell machine be compared to the salient-pole machine?

(c). How would the transient voltage differ for:

1. 0 to full load to 0 load change?
2. 0 to 200% to 0 load?
3. 0 to 300% short circuit to 0 load?

2. The paper shows the field coils located so the rotor has to be assembled around the field coils. This will be difficult to accomplish mechanically. If the coils were to be moved behind the stator coils, what would be the effect on weight, excitation, and flux leakage?

This discussion is intended to point out that there are many major inherent advantages to the salient-pole type of generator which have been the reason for its primary choice in the electrical industry almost since the birth of the industry. These very important advantages should be carefully considered before leaving the salient-pole generator for less conventional types.

References

1. Armature Reaction of Synchronous Machines, P. L. Alger. *AIEE Transactions*, vol. 47, April 1928, pp. 493–513.

J. T. Bateman: The discussers have each indicated a willingness to give thought and consideration to new electromagnetic configurations. This is gratifying since the field of rotating electric machinery design is all too often treated as a closed book.

R. T. Smith has suggested certain components of armature leakage which would be higher in the inductor-Lundell configuration than in conventional generators. Tests of the inductor-Lundell generator have demonstrated that such additional components of leakage will amount to less than half of the total armature leakage reactance. The presence of additional leakage components does not mean that increased leakage reactance is inherent in this or any other machine configuration. This leakage reactance is determined not only by machine permeances but also by the number of turns in the armature. The designer must reduce armature turns, adding to the generator weight and size if necessary, until reactances are at acceptable levels before he can compare his design with another. The discussion given in the paper assumes generators which are designed for complete compliance with aircraft application requirements and specification *MIL-G-6099A*.

K. M. Chirgwin suggests that the reduced excitation power obtained with a coaxial coil is due only to the fact that the coil is round rather than rectangular. However, in the example he has given, the increase in area of a circular core to carry three times the flux will result in an increase of the mean length of the turn of a coil surrounding that core by a factor of 1.73 rather than 3 as he states. This significant saving in excitation coil length is one of the factors which make 6-pole and 8-pole inductor-Lundell designs attractive.

W. J. Shilling has prepared a verbal analysis of the inductor-Lundell generator which presumes to prove that the configuration cannot have the weight and performance stated in the paper. Many of his discussions are inaccurate when related to the total generator design. The difficulty with such an analysis is that the preparation a long list of design problems does not in itself disprove the fact that many complete inductor-Lundell generator designs have been developed by the author and his associates where these design problems have been met effectively to provide generator sizes, weights, and performance which are attractive in the aircraft applications for which they were developed.

Dual-Mode Relay Servos

R. N. BULAND
ASSOCIATE MEMBER AIEE

N. FURUMOTO
NONMEMBER AIEE

THIS PAPER DESCRIBES a dual-mode servo which utilizes a relay or contactor for both the saturated mode and the linear mode. The dual-mode servo, which was first suggested by McDonald,[1,2] operates as an on-off or bang-bang controller for large amplitude errors (saturation mode) and as a linear servo when the error approaches zero (linear mode). The saturation mode may be optimized for a minimum transient response to step function inputs by means of special nonlinear circuitry which is used to determine the optimum points for control reversal, normally referred to as a switching line. The linear mode is optimized by inserting linear networks of the lead-lag variety.

The practical difficulty with dual-mode systems lies in combining the two modes while maintaining a reasonably simple and reliable design. Hopkins[3] suggested a saturated amplifier for the controller and analog-computer-type function generators for the nonlinear switching line. The relay is not usually favored because of the difficulty of obtaining reasonably linear action from such a nonlinear device. But the relay is the natural element for the saturation mode so that it is well worth the effort to explore the possibilities of a quasi-linear relay power amplifier.

The problem of combining a saturation mode switching line and a linear mode lead-lag network is also of great practical interest. Nieswander and MacNeal[4] develop the general design philosophy for combining both modes but design details are not considered. In this paper the design philosophy will be made more explicit for a particular type of servo and possibilities of combining the saturation mode switching line with the linear mode compensating network will be studied.

Paper 59-840, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle Wash., June 21–26, 1959. Manuscript submitted January 28, 1959; made available for printing June 2, 1959.

R. N. Buland is with the Acronutronic Systems, Inc., Glendale, Calif. N. Furumoto is with the University of California, Berkeley, Calif.

The authors wish to acknowledge the introduction of a method of linearizing relay servo systems by Dr. H. A. Wagner, former president of the H. A. Wagner Engineering Company, and now a professor at the Technical University, Aachen, Germany, and by T. F. Sturm, Vice-President of the Electronic Systems Development Corporation, Ventura, Calif.

Background

The relay has been used as a control element in regulators and servos for a least 30 years.[5] Probably there are more contactor controlled regulators and servos in existence than any other type. Yet it was not until the concept of optimum step function response was suggested that any real interest was shown in the technical literature.

In terms of optimum step function response, the relay servo is ideal because sudden reversals of control are required. On the other hand, the relay has the annoying habit of limit cycling when no error is present and the relay cannot follow arbitrary inputs accurately.

Shortly after World War II a method of linearizing relay servo systems was introduced at the Pacific Missile Range, then called Naval Air Missile Test Center, by Dr. H. A. Wagner and Mr. T. F. Sturm. This concept utilized feedback directly around the relay as shown in Fig. 1 in comparison to the conventional bang-bang relay servo of Fig. 2. The configuration of Fig. 1 was designed by linear theory assuming the relay was an infinite gain power amplifier. Bass[6] developed

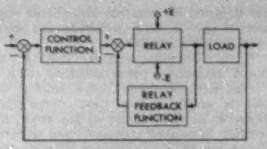


Fig. 1. Relay servo with feedback around relay

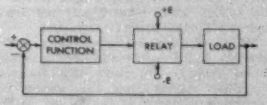


Fig. 2. Conventional bang-bang relay servo

several theorems to describe relay amplifier operation and suggested the possibility of dual-mode operation. More recently Gibson and Tuteur[7] analyzed a slightly different version of the relay amplifier with unity feedback around the relay.

The basis of linearization of the relay amplifier lies in defining the relay input to output relationship as a variable gain μ_s in a manner similar to describing function analysis.[8] It will be shown that when the relay amplifier feedback is not saturated μ_s approaches infinity and can be treated as a quasi-linear high-gain amplifier.

This paper will discuss the quasi-linear relay servo and will extend this concept to dual-mode operation.

The Dual-Mode Relay Servo

The concept of dual-mode control with the configuration of Fig. 1 is that the feedback around the relay will make the system operate in a quasi-linear fashion for small error signals. For large error signals the system will act as a conventional bang-bang controller and near optimum response can be obtained by means of a switching line in the phase plane. The concept is shown in Fig. 3. Fig. 3(A) shows the generalized block diagram where

$K_\alpha G_\alpha$ = saturated mode switching network
$K_\beta G_\beta$ = linear mode stabilizing network
K_F = coefficient relating the magnitude of the force, or torque, on the load to the magnitude of the relay output
$K_M G_M$ = dynamics of unalterable load

For small motions it will be shown that the transfer function for the system approaches the linear system shown in Fig. 3(B). For large motions the relay feedback becomes saturated and the system acts as a conventional bang-bang system as shown in Fig. 3(C). Each of the two modes will be discussed separately.

Quasi-Linear Mode

The quasi-linear mode depends upon the extent to which the relay power amplifier of Fig. 4 is linearized. For practical implementation of the relay power amplifier, a voltage amplifier with gain $-K_\beta$ is inserted as a summing point and to decrease the effective system dead space. The amount of amplification required is a function of the type, i.e., 2- or 3-position polarized relay, and quality of the relay being used and the positioning accuracy required of the system. In this study, only the 2-position relay with hysteresis will be analyzed.

The quasi-linear operation of the relay is based upon the assumption that the relay may be replaced by a gain μ_s such that

$$\mu_s(t) = \frac{E_R(t)}{e(t)}$$

where $e(t)$ and $E_R(t)$ are the instantaneous input and output of the relay. Thus μ_s is an inverse function of e.

Referring to Fig. 4(A), the relationship of E_R to α is found to be

$$\frac{E_R}{\alpha} = \frac{-1}{\frac{1}{\mu_s K_R} + K_\beta G_\beta(s)} \tag{1}$$

Because in the quasi-linear mode the error signal e is always small, it can be assumed that $1/\mu_s K_R$ is negligible, resulting in

$$\lim_{e \to 0} \frac{E_R}{\alpha} = \frac{-1}{K_\beta G_\beta(s)} \tag{2}$$

The relay power amplifier thus acts in a linear manner for small error signals. As a result, the relay with its feedback network can be replaced by a linear transfer function which is merely the negative reciprocal of the transfer function of the feedback loop.

As a partial verification of equation 2 it can be shown that the average steady-state output for the 2-position relay servo is identical to a linear system with the same $K_\beta G_\beta$. For a 3-position relay the results are identical except for a small offset caused by relay dead space. The particular case to be investigated is for $G(\beta) = (1+\tau_\beta s)^{-1}$. The steady-state response of the relay power amplifier to a step of α will be determined by equation 2 and by a detailed investigation of the relay output signal.

From the preceding equation, the steady-state response to a step input for a linear system and $\mu_s \to \infty$ is:

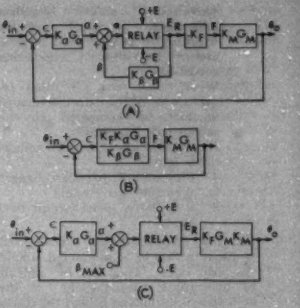


Fig. 3. A—Dual-mode relay servo. B—Quasi-linear mode, small motions. C—Saturated mode, large motions

$$E_{R_{ss}} = \lim_{s \to 0} s \cdot \frac{(1+\tau_\beta s)}{K_\beta} \cdot \frac{|\alpha|}{s} = -\frac{|\alpha|}{K_\beta} \tag{3}$$

In the Appendix, a detailed investigation of the voltages of the relay power amplifier yields the following two voltage relationships:

$$\left(EK_\beta + \frac{\Delta h}{2} + \alpha\right)\left(1 - e^{-\Delta t_1/\tau}\right) = \Delta h$$

$$\left(EK_\beta + \frac{\Delta h}{2} - \alpha\right)\left(1 - e^{-\Delta t_2/\tau}\right) = \Delta h \tag{4}$$

where Δh is the width of the hysteresis zone as shown in Fig. 4(B).

If the average output voltage on the relay arm is defined as

$$E_{R_{ss}} = \frac{\Delta t_2 E - \Delta t_1 E}{\Delta t_1 + \Delta t_2} \tag{5}$$

and the values for Δt_1 and Δt_2 from equation 4 are substituted in equation 5, $E_{R_{ss}}$ becomes

$$E_{R_{ss}} = \frac{E\left[\ln\left(1 - \frac{\Delta h}{EK_\beta + \frac{\Delta h}{2} + \alpha}\right) - \ln\left(1 - \frac{\Delta h}{EK_\beta + \frac{\Delta h}{2} - \alpha}\right)\right]}{\ln\left(1 - \frac{\Delta h}{EK_\beta + \frac{\Delta h}{2} + \alpha}\right) + \ln\left(1 - \frac{\Delta h}{EK_\beta + \frac{\Delta h}{2} - \alpha}\right)} \tag{6}$$

Now, $\Delta h << EK_\beta$ and $\ln(1+x)$ can be represented by a power series

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots, \quad -1 < x < +1$$

If $x << 1$, then the higher order terms can be neglected and $E_{R_{ss}}$ becomes

$$E_{ss_0} = \frac{-1}{K_\beta + \frac{\Delta h/2}{E}} |\alpha|$$

Since $\Delta h << E$,

$$E_{ss_0} = -\frac{1}{K_\beta} |\alpha| \tag{7}$$

as previously obtained, i.e., the steady response to a step input of α is the same as the average response to a constant input of α.

From equation 6 it can be seen that the system is linear in the steady-state case as long as the following relationship holds

$$|\alpha| \ll \left| EK_\beta \right| - \left| \frac{\Delta h}{2} \right| \tag{8}$$

When $|\alpha| = |EK_\beta| - |\Delta h/2|$, the system is in the saturated mode.

At steady-state conditions the relay amplifier tends to oscillate about the hysteresis zone. The period and amplitude of the oscillation will depend upon the parameters of the system. The gain, μ_s, can be found for this condition to be

$$\mu_s = \frac{\tau E K_\epsilon}{\Delta h}$$

For a practical situation μ_s will be in the order of 10,000.

Saturated Mode

The saturated mode will be designed to obtain near optimum response to step function inputs. In geneal the number of reversals of control for an optimum saturated response will be $N-1$ where N is the order of the differential equation describing the load.[9] However, Chang[10] utilized the dead space of a 3-position relay control to allow the higher order transients to die out during the switching sequence and effectively reduce the system to a second order system so that only one switching curve was required.

The procedure in this paper will be to design a single switching line which will approximate the first switching curve for the particular load function under consideration. The system will be designed to have a linear region about the switching line and about the zero error or terminal point. This linear region will be relied upon to compensate for any other switching curves that may be required by the criteria of reference 9 and for the fact that a switching line will be utilized rather than the actual switching curve. In other words, a lead network will be used to approximate the optimum switching curve.

For a second order system conventional phase plane techniques can be employed to determine the optimum switching curve[11] as shown in Fig. 5. The switching curve may be approximated by a straight line but then an optimum response is obtained for only one value of step input amplitude. The quasi-linear mode is relied upon to minimize the error in switching. The instrumentation of the straight line is, of course, much simpler than the curve for the switching line.

$$[\epsilon + T_L \dot{\epsilon}] = \epsilon[1 + T_L s]$$

The switching line circuit can be inserted in the feedback or the forward path of the servo. The forward path is the more difficult problem because the switching circuit must be approximated by a lead-lag network

$$\frac{1 + T_L s}{1 + \tau s}.$$

The additional lag makes the system third order, and two switchings are required for an optimum response. Again the quasi-linear zone is utilized to minimize this additional error.

For servos of higher order the single switching line is still useful but the width of the quasi-linear region becomes of increasing importance and the transient time will not approach the optimum as closely.

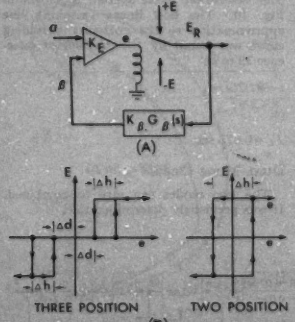


Fig. 4. A—Relay power amplifier. B—Relay characteristics

Basic Design

The design of the dual-mode relay servo can be described, with the aid of Fig. 3, in the following steps:

1. The quasi-linear mode is designed by conventional linear techniques so that

$$\frac{\theta_0}{\epsilon} = K_L K_M G_L G_M \tag{9}$$

$$\frac{\theta_0}{\theta_{in}} = \frac{K_L K_M G_L G_M}{1 + K_L K_M G_L G_M} \tag{10}$$

where

$$K_L = \frac{K_a K_F}{K_\beta} \tag{11}$$

$$G_L = \frac{G_a}{G_\beta} \tag{12}$$

2. The boundary between the saturation mode and the quasi-linear mode is determined when

$$\beta_{max} = \alpha$$

but

$$\beta_{max} = K_\beta E$$

and

$$\alpha = K_{a_1} \epsilon$$

after all transients have died out so that all derivatives of ϵ are zero then

$$|\epsilon_{LM}| = \frac{K_\beta}{K_{a_1}} |E| \tag{13}$$

where

ϵ_{LM} = the minimum saturated mode error

E = maximum magnitude of E_R

as shown by equation 8. The linear limit will depend upon the quality of the relay, character of G_a and G_β and the derivatives of ϵ.

For preliminary design ϵ_{LM} will be utilized as a measure of the quasi-linear region. From equation 13 it is seen that

$$\epsilon_{LM} = f(E)$$

but E is also constrained by the output response

$$\theta_0 = K_M G_M K_F E_R \tag{14}$$

Thus the speed of response is directly related to the magnitude of K_F and E_R or E for the saturated mode. Normally it is desirable to make K_F and E as large as possible in order to minimize response time. Also K_β and K_{a_1} are constrained by the quasi-linear gain K_L and by K_F. Combining equations 11 and 13

$$\frac{\epsilon_{LM}}{E} = \frac{K_F}{K_L} \tag{15}$$

which defines the fraction of the linear range

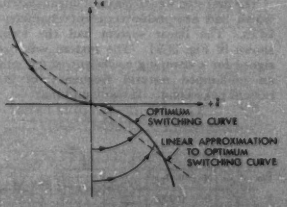


Fig. 5. A phase-plane plot of the optimum switching curve for a second-order bang-bang servo system

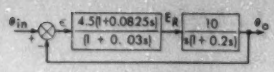


Fig. 6. Linear design

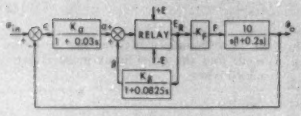


Fig. 7. Quasi-linear relay servo

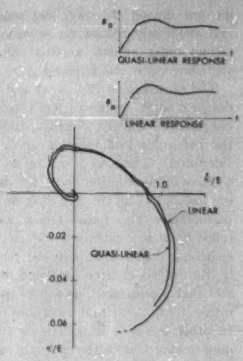


Fig. 8 (right). Comparison of quasi-linear and linear systems for small magnitude inputs ($\theta_{in} = 0.3\epsilon_{LM}$)

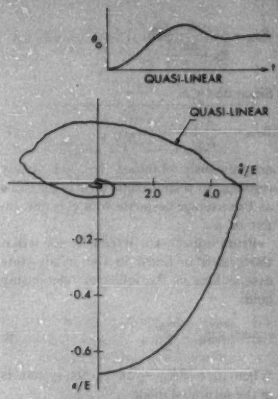


Fig. 9. Quasi-linear system response for large magnitude inputs ($\theta_{in} = 3\epsilon_{LM}$)

with respect to the maximum relay output.

3. The saturation mode switching line is defined by G_α and will normally have the form

$$G_\alpha = \frac{1+T_L s}{1+\tau_1 s} \tag{16}$$

where τ_1 is chosen to be a part of G_L if this is possible. T_L is chosen to be a best fit to the optimum first switching line.

4. The final step is to choose G_β from equations 12 and 16

$$G_\beta = \frac{G_\alpha}{G_L} \tag{17}$$

and choosing either K_β or K_α arbitrarily so that the other can be found from equation 11. As a final step the analysis should be reviewed to determine where simplifications can be made.

Example Problem

To demonstrate the dual-mode relay servo design, a second order servo was designed and tested by means of an analog computer. The unalterable element of the system was

$$K_M G_M = \frac{10}{s(1+0.2s)}$$

The Quasi-Linear Mode

The first step was to design a linear system which had acceptable transient characteristics. The linear system had the form shown in Fig. 3(B). The system was designed for a damping coefficient of 0.5 and an undamped natural frequency of 20 radians per second. It was found that

$$T_L = 0.0825$$

$$\tau_1 = 0.03$$

$$K_L = 4.5$$

and

$$\frac{\theta_o}{\epsilon} = \frac{45.0(1+0.0825s)}{s(1+0.2s)(1+0.03s)}$$

as shown in Fig. 6. The quasi-linear relay is shown in Fig. 7 where

$$K_L = 4.5$$

$$G_L = \frac{1+0.0825s}{1+0.03s}$$

and

$$G_\alpha = \frac{1}{1+0.03s}$$

$$G_\beta = \frac{1}{1+0.0825s}$$

The next step in the design is to determine the magnitudes of K_F and E. For the system being designed, which was built up on an analog computer, the maximum E was ± 50 volts, the maximum K_F is ± 1. If these values are used, then

$$\frac{\epsilon_{LM}}{50} = \frac{1}{4.5} = 0.22$$

or ϵ_{LM} is 22% of E. This is a quite acceptable linear zone and has been found to operate quite satisfactorily although Hopkin, in Reference 3, recommended a minimum linear region in order to sharpen the saturated mode response.

For comparison purposes K_α was set at 4.5 and thus

$$K_\beta = 1.0$$

This system was then compared to the equivalent linear system. In Fig. 8 the two systems are compared for a step input

$$\theta_{in} = 0.3\epsilon_{LM}$$

The response of the quasi-linear system is shown for

$$\theta_{in} = 3\epsilon_{LM}$$

in Fig. 9. The response is by no means optimum although the per cent overshoot is about the same as for the linear model.

Saturated Mode Design

The saturation mode switching line must be computed to determine G_α. From the work done on the quasi-linear system it seems that a suitable form for G_α would be

$$G_\alpha = \frac{1+T_L s}{1+0.03s}$$

The saturation mode system is shown in Fig. 10. The best linear switching line approximation to the optimum switching curve was found by the method of reference 12 to be

$$\epsilon + 0.075\dot{\epsilon} = 0$$

or

$$T_L = 0.075 \text{ sec.}$$

Dual-Mode Design

The two modes can now be combined. It was previously determined that

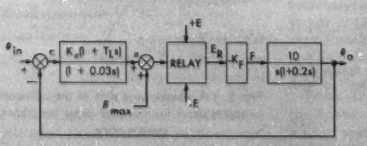


Fig. 10 (left). Saturation mode

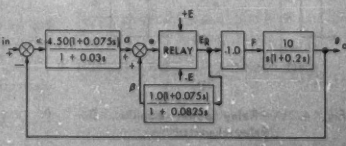


Fig. 11 (right). Final design, Case 1

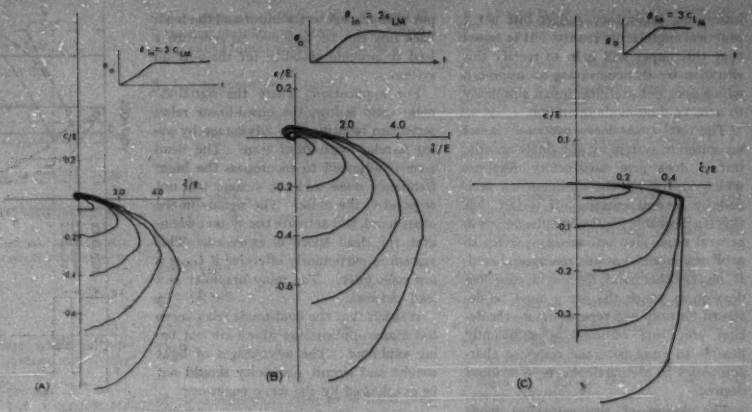


Fig. 12. Family of responses for dual-mode relay servo

A—Case 1 conditions
B—Case 2 conditions
C—Case 3 conditions

$$G_L = \frac{1+0.0825s}{1+0.03s}$$

from the quasi-linear design, and that

$$G_n = \frac{1+0.075s}{1+0.03s}$$

from the saturated mode design for the numerator and the quasi-linear design for the denominator. Thus G_β is defined as

$$G_\beta = \frac{1+0.075s}{1+0.0825s}$$

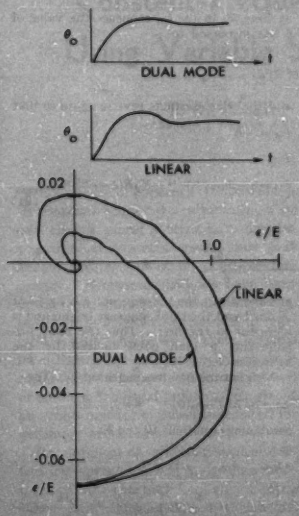


Fig. 13. Response of dual mode and linear systems to small magnitude input and Case 1 conditions

All that remains is to choose K_n and K_β, keeping the ratio fixed at 4.5. Experiments indicate that large values of K_n and K_β would be more desirable because the effects of relay dead space is reduced. However, in order to determine the effect, two values of K_n and K_β will be investigated.

Case 1. $K_\gamma = 1$ $K_n = 4.5$ $K_\beta = 1.0$
Case 2. $K_\gamma = 1$ $K_n = 0.45$ $K_\beta = 0.1$

The final design for Case 1 is shown in Fig. 11. A family of phase-plane responses to different amplitude steps are shown in Figs. 12(A) and 12(B). The transient response is shown for

$\theta_{in} = 3\epsilon_{LM}$

In order to determine the effect of a narrower linear band, Case 3 was investigated:

Case 3. $K_\gamma = 0.1$ $K_n = 4.5$ $K_\beta = 0.1$

Thus K_L is unchanged, but the ratio

$$\frac{\epsilon_{LM}}{E} = 0.022$$

or ϵ_{LM} is only 2.2% of E. The phase plane for a family of inputs is shown in Fig. 12(C). The transient response is shown for

$\theta_{in} = 3\epsilon_{LM}$

The response is much slower because the forcing function has been reduced by a factor of 10 but the switching is much sharper as would be expected, because of the smaller linear region.

The response of the dual-mode relay servo for the Case 1 gains is compared to the linear system response in Fig. 13 for a small amplitude input

$\theta_{in} = 0.3\epsilon_{LM}$

It is seen that the dual-mode response to small amplitude inputs does not compare to the linear as well as the quasi-linear response did in Fig. 8. However, the overshoot has been reduced and the over-all response time is about the same.

The frequency response of the dual mode servo with Case 1 gains was also investigated briefly. Fig. 14 shows the input and output for

$\theta_{in} = 0.3\epsilon_{LM}$

at 3 cycles per second, and

$\theta_{in} = 3\epsilon_{LM}$

at 1 cycle per second.

Discussion

The theory for dual-mode servos has been developed in the literature to a considerable extent and experimental models have been built to demonstrate the theory. Yet very little work has been devoted to developing practical circuitry for dual-mode servos. In other words, the dual-mode servo is, at present, an

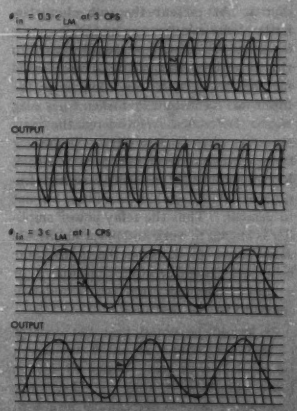


Fig. 14. Frequency response of dual-mode relay servo

interesting laboratory gadget but not a practical engineering reality. It is hoped that this paper will help to rectify this situation by demonstrating an approach which does not sacrifice design simplicity to achieve dual-mode operation.

The dual-mode servo proposed is not an optimum system by any single criteria, but it does give satisfactory response which approaches the optimum as specified by several criteria. It is felt that this is a desirable feature because servos in general must give satisfactory service to arbitrary inputs, not just one special kind. If, on the other hand, the special case does develop in which the servo must be designed for only one type of input, the design procedure outlined is sufficiently flexible to emphasize the response characteristics to the particular type of input desired.

The design of the saturated mode switching line is straightforward for a second order system but becomes a tedious computation for third order and an almost impossible problem for orders above third except by using a large scale computer. When using a computer to design the switching line for a high order system, the quasi-linear mode will modify the transient markedly and should be programmed on the computer. The width of the quasi-linear zone may be of first order importance in such a design.

The design of the quasi-linear mode is straightforward linear design. Unfortunately the deviation of the quasi-linear system from the linear is difficult to predict. Work needs to be done in this area so that the action of the relay power amplifier can be more accurately predicted. At present the gain characteristics of the relay power amplifier can be described for only a few special cases. In this paper the average gain μ_e was derived for the steady-state condition when the input was a constant and when G_β is a lag term. Reference 7 considered the same situation when G_β was unity. Describing functions can be used if the relay feedback loop is sufficiently insensitive to high frequencies, with G_β a low pass filter, but not in general. Thus the relay power amplifier is known to work and linear analysis is satisfactory to describe the quasi-linear relay servo to a first approximation. For a more detailed analysis, simulation seems to be the best tool.

It should be noted that the concept of feedback around an on-off device can be applied, with good results, to devices more complicated than a relay. For example, a pulse coder can be built with feedback around the coder. This feedback will linearize the average coder output with respect to the input and the feedback link can be designed to generate a lead to provide stability for the over-all system.

For applications where the signal-to-noise ratio is low, the quasi-linear relay servo can be utilized to advantage by using a relay with dead zone. The dead zone is adjusted to encompass the high-frequency noise and the system will not respond to the noise. The signal, on the other hand, will activate the system whenever the dead space is exceeded. This system is particularly effective if G_β is a low pass filter. The relay amplifier is a lead network.

It is felt that the dual-mode relay servo has many applications which are not being exploited. The advantages of light weight and circuit simplicity should not be overlooked by the servo engineer.

Conclusions

A design concept has been developed for a dual-mode relay servo which employs linear networks in both the saturated mode and the linear mode. The linear mode employs a quasi-linear relay amplifier. It is concluded that:

1. This design provides near optimum step response in the saturated mode and near linear response in the quasi-linear mode with no additional complexity beyond a lead lag network.
2. Linear theory can be used in designing the quasi-linear relay servo with very good results.
3. The relay amplifier is an excellent device for obtaining high power amplification without sacrificing the simplicity of the relay or the advantages of linear design techniques.

Appendix. Development of Quasi-Linear Relay Amplifier Equations

The relay amplifier is shown in Fig. 4(A). For the 2-position relay case the feedback voltage β is approaching either $+$ or $-K_\beta E$. However e is a function of β and limits β to a range of values determined by the hysteresis zone, Δh, of the relay.

Consider the case when

$$G_\beta = \frac{1}{1+\tau_\beta s}$$

and α is a step function. The history of e and E_R are shown in Fig. 15. The analysis will be piecewise linear beginning a new time sequence with each switching of E.

For the $t_{(0)}$ time sequence e initial equals α and

$$E_R = -E$$

Then

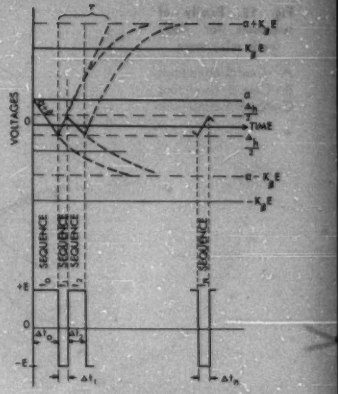


Fig. 15. Steady-state response of the 2-position relay power amplifier to a step input

$$e = \alpha - K_\beta E(1 - e^{-t(0)/\tau_\beta}) \tag{18}$$

when e equals $-\Delta h/2$ the relay contacts reverse and the $t_{(1)}$ sequence beings

$$e_{(1)} = \alpha - \beta_{(0)} e^{-t_{(1)}/\tau_\beta} + K_\beta E(1 - e^{-t_{(1)}/\tau_\beta}) \tag{19}$$

where

$$\beta_{(0)} = \alpha + \frac{\Delta h}{2} \tag{20}$$

at time Δt_1 in the $t_{(1)}$ sequence the value of e has increased to

$$e = \frac{\Delta h}{2}$$

and the relay contacts reverse again so that

$$E_R = -E$$

and

$$e_{(2)} = \alpha - \beta_{(1)} e^{-t_{(1)}/\tau_\beta} - K_\beta E(1 - e^{-t_{(1)}/\tau_\beta}) \tag{21}$$

where

$$\beta_{(1)} = \alpha - \frac{\Delta h}{2} \tag{22}$$

at time Δt_2 in the $t_{(2)}$ sequence e is again at $-\Delta h/2$ and the third sequence begins and is identical to the first. Thus the relay has gone into a limit cycle limited by the hysteresis band.

Now in time Δt_1 e has increased by Δh or

$$\Delta h = e_n(\Delta t_1) - e_n(0) \qquad n = 1, 2, 3, \ldots$$

combining equations 19 and 20

$$\Delta h = \alpha - \left(\alpha + \frac{\Delta h}{2}\right) e^{-\Delta t_1/\tau_\beta} + K_\beta E(1 - e^{-\Delta t_1/\tau_\beta}) + \frac{\Delta h}{2} \tag{23}$$

$$\Delta h = \left(\frac{\Delta h}{2} + K_\beta E + \alpha\right)(1 - e^{-\Delta t_1/\tau_\beta})$$

Also

$$-\Delta h = \alpha_m(\Delta t_1) - \alpha_m(0) \qquad n = 1, 2, 3, 4, \ldots$$

combining equations 21 and 22

$$-\Delta h = \alpha - \left(\alpha - \frac{\Delta h}{2}\right) e^{-\Delta t_2/\tau_2} - K_p E(1 - e^{-\Delta t_2/\tau_2})$$

$$\Delta h = \left(\frac{\Delta h}{2} + K_p E - \alpha\right)(1 - e^{-\Delta t_2/\tau_2}) \qquad (24)$$

Equations 23 and 24 are equation 4 of the text.

It should be noted that Δt_1 and Δt_2 are almost equal when α is small compared to $K_p E$ but when

$$\alpha = K_p E - \frac{\Delta h}{2}$$

then Δt_1 approaches zero while Δt_2 approaches ∞ and the system enters the saturated zone.

The analysis for the 3-position relay servo follows the same form but is complicated by the dead zone, the configuration of the G_p network, and whether or not the relay is grounded in the dead zone position. The 3-position relay will always have a positioning error related to one half the dead space of the relay.

For the 2-position relay when α is small compared to $K_p E$ the relay error voltage is essentially a sawtooth wave while E_R is a square wave. The period of the oscillation is determined by τ_2, K_p, K_s, and E. If the first harmonic of E_R and e are utilized, a describing function can be generated.

$$E_R(\text{1st harmonic}) = \frac{4E}{\pi} \cos \omega t$$

$$e(\text{1st harmonic}) = \frac{4}{\pi 2} \frac{\Delta h}{K_s} \cos \omega t$$

Then

$$\mu_s = \frac{E_R}{e} = \frac{\pi E K_s}{\Delta h}$$

For some typical values let

$$E = 50 \text{ volts}$$

$$\Delta h = 0.5 \text{ volt}$$

$$K_s = 25$$

Then

$$\mu_s = 7{,}800$$

which is sufficiently close to ∞ for a practical case.

References

1. Nonlinear Techniques for Improving Servo Performance. D. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 400–31.

2. Multiple Mode Operation of Servomechanisms. D. McDonald. *Bulletin 3-3*, Cook Research Laboratories, Chicago, Ill., 1951.

3. A Phase-Plane Approach to the Compensation of Saturating Servomechanisms. A. M. Hopkin. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 631–39.

4. Optimization of Nonlinear Control Systems by Means of Nonlinear Feedbacks. R. S. Nieswander, R. H. MacNeal. *Ibid.*, pt. II (*Applications and Industry*), vol. 72, Sept. 1953, pp. 262–72.

5. Automatic Voltage Regulators, Application to Power Transmission Systems. C. A. Nickle, R. M. Carothers. *Ibid.*, vol. 47, July 1928, pp. 967–74.

6. Equivalent Linearization, Nonlinear Circuit Synthesis, and the Stabilization and Optimization of Control Systems. R. W. Bass. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Brooklyn Polytechnic Institute, New York, N. Y., Oct. 1956.

7. The Response of Relay Amplifiers with Feedback. J. E. Gibson, F. B. Tuteur. *AIEE Transactions*, vol. 76, pt. II, Nov. 1957, pp. 303–07.

8. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms. E. J. Kochenburger. *Ibid.*, vol. 69, pt. I, 1950, pp. 270–84.

9. An Investigation of the Switching Criteria for High Order Contactor Servomechanisms. I. Bogner, L. F. Kazda. *Ibid.*, vol. 73, pt. II, July 1954, pp. 118–27.

10. Optimum Switching Criteria for Higher Order Contactor Servo with Interrupted Circuits. S. S. L. Chang. *Ibid.*, vol. 74, pt. II, Nov. 1955, pp. 272–76.

11. Differential Equations with a Discontinuous Forcing Term. D. W. Bushaw. *Report no. 469*, "Experimental Towing Tank," Stevens Institute of Technology, Hoboken, N. J., Jan. 1953.

12. Optimum Switching of Third Order Position Relay Servos for Step Inputs. R. N. Buland. *M. S. Thesis*, University of California, Berkeley, Calif., June 1955.

Constant-Frequency A-C Power Using Variable Speed Generation

R. D. JESSEE
ASSOCIATE MEMBER AIEE

W. J. SPAVEN
NONMEMBER AIEE

THE USUAL approach to provide a constant-frequency alternating-current electric power system is to drive a generator at constant speed, thus producing output power at constant frequency. Numerous methods have been devised for obtaining constant frequency power from a generator running at variable speeds. Analysis of these systems have shown the vast majority of these methods to be impractical for one reason or another, the usual problem being that an excess amount of power is required for generator excitation.

Development of switching semiconductors has opened the way to new methods for obtaining constant frequency electric power from variable speed generators. In the unique system which is described, the generator produces voltage which varies in frequency with the generator speed. The generator output power is directed into an efficient switching converter which changes the form of the voltage, resulting in output voltage of a lower frequency. By varying the rate of switching in the converter as the generator speed changes, the frequency of the output voltage from the converter may be kept constant within desired limits. Voltage regulation may be accomplished in the usual manner, by controlling the generator field current.

Variable-Speed Constant-Frequency System

The variable-speed constant-frequency electric power system is shown in block diagram form in Fig. 1. The generator develops substantially a sine wave voltage at a frequency proportional to the generator speed. The minimum operating frequency of the generator is chosen to be several times higher than the desired system frequency. The electric power developed by the generator is transferred to the switching frequency converter, where periodic switching converts the frequency of the generated voltage to a lower frequency. The fundamental frequency of the converter output voltage (and therefore the system voltage) is the difference between the generator frequency and the switching frequency of the converter. Because the converter output voltage is derived from a periodic switching operation, this voltage contains high frequency components which may be undesirable if applied to the load. For

Paper 59-810, recommended by the AIEE Rotating Machinery Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 29, 1959.

R. D. Jessee is with Westinghouse Electric Corporation, Lima, Ohio. W. J. Spaven is with Walter Kidde & Company, Inc., Bellville, N. J.

The authors wish to acknowledge the contributions of R. F. Blake, E. Demers, and C. C. Ware co-inventors of the basic frequency conversion methods. Acknowledgement is also made of independent work in a similar area by the Canadian Westinghouse Company.

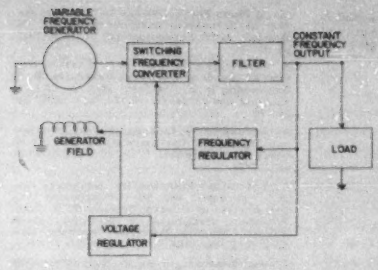


Fig. 1 (left). System diagram

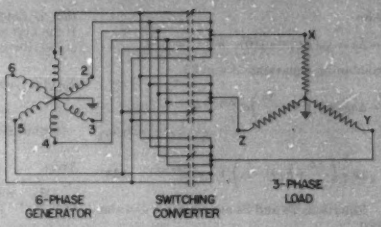


Fig. 2 (right). Schematic diagram of power circuits through switches of converter

this reason, a filter is included in the system to prevent these high frequency voltages from reaching the load.

The system frequency is kept constant by controlling the switching rate of the converter to maintain a constant difference between the switching frequency and the generator frequency despite changes in generator speed. The switching frequency is controlled by means of a frequency regulator which senses the system frequency and provides an error signal to the switching frequency converter. The converter corrects its switching frequency to reduce the error signal and thereby keep the system frequency within desired limits.

Changes in generator speed and in load cause changes in excitation requirements in the generator for constant system voltage. Voltage regulation is accomplished by a voltage regulator which senses the system voltage and adjusts the generator field current to the value required to maintain the desired system voltage.

Method of Frequency Conversion

The method of frequency conversion to be discussed has many variations. All these variations give duplicate or similar results so far as output voltage is concerned. There are, however, numerous design considerations which must not be overlooked in coosing a particular switching method for an electric power system. For the purpose of this presentation, it appears logical to choose the method which is most straightforward and understandable, even though it may not be best from the standpoint of design.

Consider the system shown in Fig. 2. This system consists of a 6-phase generator connected to a 3-phase load through an array of 18 switches. In this case, the filter is considered part of the load. The action of the switches is to connect the load to three of the six phases for a prescribed time, then to advance to the next set for an equal time interval, and again advance, etc.

For example, at some time the load point X is connected to generator terminal 1, while Y is connected to 3, and Z to 5. At the switching instant, the three switches making the connections just described open, and three others connect X to 2, Y to 4, and Z to 6. Terminal X is then connected in sequence to 3, 4, 5, 6, etc., for equal time intervals, while Y is connected to 5, 6, 1, 2, etc., and Z to 1, 2, 3, 4, etc.

The result of this switching action may be seen in Fig. 3, which shows all six phase voltages of the generator. Superimposed on these is one phase of the load voltages which is a result of switching that load from generator terminal 1 to 2 to 3 to 4, etc. It can be seen that this load voltage is predominantly a sine wave with high frequencies superimposed on it. The fundamental component of the load voltage is drawn into the figure for clarity. The frequency of the load voltage is determined by the frequency of the generator voltage and the switching frequency. It is shown in Appendix I that the load voltage frequency is the difference between generator and switching frequencies.

In order to make it legible, Fig. 3 shows only phase one of the three load voltages. However, the other two phases can be drawn in and their fundamentals will be displaced 120 and 240 degrees from the wave form shown.

The voltage appearing on one phase of the load is derived in Appendix I, equation 23 and is expressed as

$$V_x = \frac{3E}{\pi}\left[\cos(\alpha-\beta+60°) - 1/5\cos(5\alpha+\beta-60°) + 1/7\cos(7\alpha-\beta+60°) - \ldots\right]$$

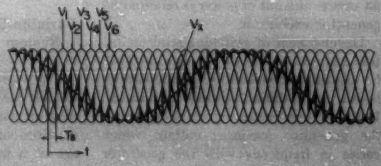


Fig. 3. Voltage waveforms of circuit on Fig. 2, showing output voltage superimposed on six generated voltages

where α is proportional to the switching frequency, β is proportional to the generator frequency, and E equals the maximum value of the generator voltage. Inspection of this equation shows that the fundamental frequency of the load voltage is equal to the difference between the generator frequency and the switching frequency. Therefore, if the generator frequency changes, the output frequency may be kept constant by maintaining a constant difference between the generator frequency and the switching frequency.

This load voltage equation also offers valuable information about the undesirable high-frequency voltages. The lowest frequency undesirable voltage is defined by either the second or third term of the equation, depending on the generator and switching frequencies. For example, suppose the desired output frequency is 400 cycles per second (cps), from a 1,600-cps generator. Since the output frequency is the difference between generator and switching frequencies, the switching frequency, f_s, can be either 2,000 cps or 1,200 cps. If $f_s = 2{,}000$ cps, the second and third terms of the equation yield $5(2{,}000)+1{,}600 = 11{,}600$ cps, and $7(2{,}000)-1{,}600 = 12{,}400$ cps respectively. The lowest undesirable frequency is $(11{,}600/400 = 29)$ times the fundamental. If $f_s = 1{,}200$ cps, the second term yields $5(1{,}200)+1{,}600 = 7{,}600$ cps, and the third term $7(1{,}200)-1{,}600 = 6{,}800$ cps. The lowest undesirable frequency here is $(6{,}800/400 = 17)$ times the fundamental.

The figures in the example indicate that for a given generator design it is desirable to switch at a frequency higher than that of the generator because this results in higher minimum undesirable frequencies. Of course, the higher the frequency of the

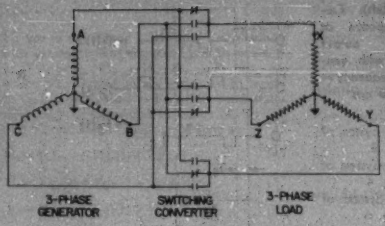


Fig. 4 (left). Schematic diagram of power circuits through switches of converter (simplified system)

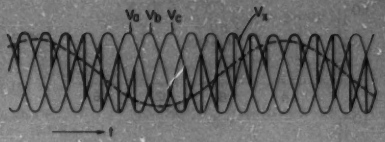


Fig. 5. Voltage waveforms of circuit of Fig. 4, showing output voltage superimposed on three generated voltages

undesirable voltage the easier will be the problem of filtering it. Another important consideration is that both the generator frequency and the switching frequency be made as high as practical in order to keep the filtering problem to a minimum.

In Appendix I the 3-phase output voltages are given in equations 23, 24, and 25. The fundamental frequency component of these voltages is given in the first term of each equation.

$$V_{xf} = \frac{3E}{\pi} \cos(\alpha - \beta + 60°)$$

$$V_{yf} = \frac{3E}{\pi} \cos(\alpha - \beta + 180°)$$

$$V_{zf} = \frac{3E}{\pi} \cos(\alpha - \beta - 60°)$$

It will be noted from these equations that each of the three voltages is displaced 120 degrees from the other two. In this respect, the output voltages are normal for a 3-phase generating system.

Again referring to the equations of the fundamental frequency voltages, assume that the switching frequency is higher than the generator frequency. This results in the value of $(\alpha - \beta)$ being positive. In this case, V_x is 60 degrees ahead of the reference, V_y is 180 degrees ahead of the reference, and V_z is 60 degrees behind the reference. Therefore, the sequence of the output voltages is V_x, V_y, V_z. If, however, the generator frequency be higher than the switching frequency, the value of $(\alpha - \beta)$ is negative; and V_x is 60 degrees behind the reference, V_y is 180 degrees behind the reference, while V_z is 60 degrees ahead of the reference. The sequence of this mode of operation is V_x, V_y, V_z, opposite to that of the previous case. This is an important characteristic in some applications of this system. Appendix III describes one such application.

Switching Frequency Converter Variations

It has already been stated that there are many possible variations on the basic system described. One which provides simplification in the switching circuits should be considered and is shown in Fig. 4. The switch operation is the same as for the previous case with the exception that there are only half as many switches in this system. Load point X is switched from V_a to V_b to V_c, etc. at equal time intervals. It can be shown that the voltage appearing on the load may be expressed as

$$V_x = \frac{3\sqrt{3}E}{2\pi} [\cos(\alpha - \beta - 60°) + 1/2 \cos(2\alpha + \beta - 120°) + 1/4 \cos(4\alpha - \beta - 60°) + 1/5 \cos(5\alpha + \beta - 120°) + 1/7 \cos(7\alpha - \beta - 60°) + 1/8 \cos(8\alpha + \beta - 120°) + \ldots]$$

The load voltages V_y and V_z are similar to V_x but displaced 120 and 240 degrees respectively. Inspection of the equation of load voltage for this case reveals the load voltage contains all the frequencies present in the 6-phase generator system. In addition, lower frequency voltage components are present as seen in the second and third terms of the equation. Because of these lower frequency voltage components, this system requires more filtering for comparable quality output. Therefore, the system requirements must be considered before the switching method is fixed. For some applications, the switching method of Fig. 4 is advantageous; for example, where sine wave output is not a requirement, or where a large number of switches is more objectionable than is the resulting increase in filter weight. The unfiltered output voltage waveform is shown in Fig. 5.

The simplest form that this method of frequency conversion can take is shown in Fig. 6 where a single-phase voltage is switched to the load in alternating polarity. Here again the load voltage contains a component of voltage which has a frequency equal to the difference between generator frequency and switching frequency. This voltage may be expressed as

$$V_L = \frac{2E}{\pi} [\cos(\alpha - \beta) + \cos(\alpha + \beta) - 1/3 \cos(3\alpha - \beta) - 1/3 \cos(3\alpha + \beta) + 1/5 \cos(5\alpha - \beta) + 1/5 \cos(5\alpha + \beta) + \ldots]$$

Inspection of this equation shows that, in addition to the fundamental frequency component, there exists a voltage component of equal magnitude at a relatively low frequency; the sum of the generator and the switching frequencies. This is, of course, quite objectionable because this component is extremely difficult to filter out in a practical power system. Fig. 7 shows the waveform of the load voltage.

Fig. 8 shows a comparison of the frequencies present in the output voltages of the three different switching methods shown in Figs. 2, 4, and 6. Figures 8(A) and 8(B) show the relative magnitudes of the various voltage components and their frequencies for the system using the 6-phase generator. Figs. 8(C) and 8(D) show the same for the 3-phase generator switched with nine switches. Figs. 8(E) and 8(F) show the same data for the single-phase case. From this compari-

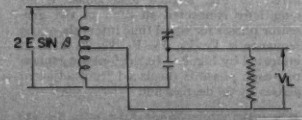


Fig. 6. Schematic diagram of basic single-phase switching circuit

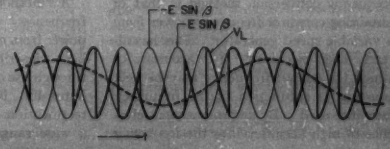


Fig. 7. Voltage waveforms of circuit in Fig. 6

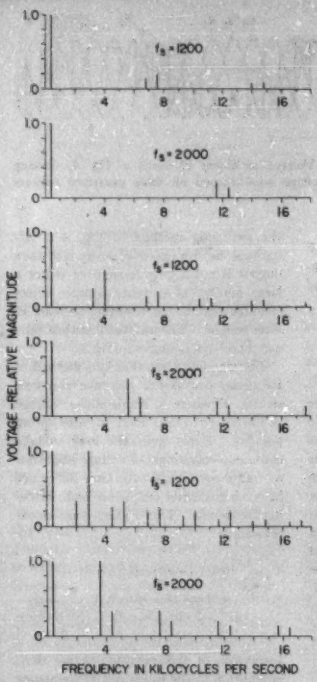


Fig. 8 (left). Frequency spectra of converter output voltages with generator frequency = 1,600 cps

A and B—System of Fig. 2
C and D—System of Fig. 4
E and F—System of Fig. 6

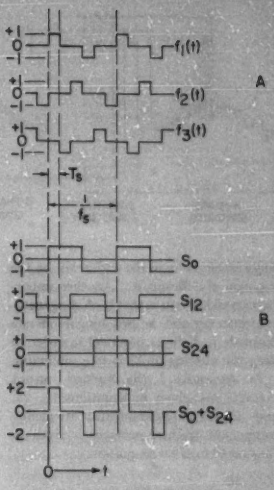


Fig. 9 (right). Time varying functions describing converter switching

A—Switching functions
B—Derivation of switching functions from basic square waves

son it is seen that the 6-phase switching method gives the best wave form of the three. The wave form could be further improved by switching more phases, but the obvious objection is that more switches are required—an undesirable feature from the standpoint of cost and reliability.

Switching Semiconductors

Several methods are available for using semiconductors as switches. Some of these methods and semiconductors used are discussed briefly in Appendix II.

Application

The most obvious application for this frequency conversion method is that of obtaining constant frequency power from a variable speed generator. This has already been discussed. The accuracy to which the output frequency may be held constant is a function of the frequency reference used. By incorporating a reference of high accuracy into the frequency regulator, an electric power system having precise frequency may be built. Fig. 12 gives the estimated weight per kva rating for this type of system.

Appendix III describes a system which is an important application of this method of frequency conversion. In this system, the frequency converter is used to supply controlled frequency excitation power to the generator. The power required from the frequency converter depends on the speed range of the generator. In a system having a narrow speed range, precise frequency output power in relatively large ratings may be obtained with a frequency converter having a low rating.

Conclusions

A method for obtaining constant-frequency alternating-current electric power from a variable speed generator has been described. This is accomplished through use of switching circuits which are naturally bilateral, without previous rectification. This allows operation over a wide range of load power factors because there is always a low-impedance path between the load and the generator regardless of the direction of current flow.

One of the major advantages in using this system is that the only moving part required is the generator rotor. The absence of mechanical devices is indicative of lower maintenance requirements.

Because the system frequency is controlled by electrical means much faster than mechanical control, frequency transients caused by changes in generator speed are quite small and very short in duration. This is an important consideration in many applications.

By use of semiconductors which are presently available, it is practical to build electric systems in small ratings up to several kilovolt-amperes having precise frequency control. As the development of higher power semiconductors progresses, much higher power ratings will become feasible.

Appendix I. Output Voltage Calculations

The output voltage to the load may be expressed mathematically and is developed here. In Fig. 2 each load phase is sequentially connected to each of the generator phases for equal time intervals. The sequence is V_1, V_2, V_3, V_4, V_5, V_6, etc., as indicated by Fig. 3. In the 6-phase generator, each phase is displaced from adjacent phases by 60 degrees and

$$V_1 = -V_4 \quad (1)$$

$$V_2 = -V_5 \quad (2)$$

$$V_3 = -V_4 \tag{3}$$

By use of equations 1, 2, and 3 it is evident that an equivalent sequence can be written as V_1, $-V_5$, V_2, $-V_6$, V_3, $-V_4$ etc., in which V_5, V_6, and V_4 are expressed in terms of $-V_2$, $-V_1$, and $-V_3$. The load voltages may be expressed in terms of V_1, V_2, and V_3 and three functions of time as

$$V_x = V_1 f_1(t) + V_2 f_2(t) + V_3 f_3(t) \tag{4}$$

$$V_y = V_1 f_2(t) + V_2 f_3(t) + V_3 f_1(t) \tag{5}$$

$$V_z = V_1 f_3(t) + V_2 f_1(t) + V_3 f_2(t) \tag{6}$$

where $f_1(t)$, $f_2(t)$, and $f_3(t)$ are the required functions of time shown in Fig. 9(A). The value of $f_1(t)$ is $(+1)$ for the first time interval (from $t=0$ to $t=T_2$), zero for the next two intervals, (-1) for the fourth, zero for the following two, etc. The values of $f_2(t)$ and $f_3(t)$ follow a similar pattern, but are shifted in time as shown in Fig. 9(A). By inspection of Fig. 9(A) it is seen that equation 4 has values of V_x equal to V_1 during the first interval, $-V_2$ during the second, V_3 during the third, etc., resulting in the required sequence. The values of V_y and V_z follow the same sequence but are displaced in time.

It is more convenient to express $f_1(t)$, $f_2(t)$, and $f_3(t)$ in terms of square-wave time functions. Fig. 9(B) shows three unit square waves, S_0, S_{12}, and S_{24}: displaced 120 degrees from each other. It is also shown that the sum of two square waves (S_0+S_{24}) results in the waveform of $f_1(t)$, but with twice the magnitude. Therefore,

$$f_1(t) = (1/2)(S_0 + S_{24}) \tag{7}$$

$$f_2(t) = (1/2)(S_{12} + S_0) \tag{8}$$

$$f_3(t) = (1/2)(S_{24} + S_{12}) \tag{9}$$

Equations 1, 2, and 3 may now be written in terms of the square wave functions.

$$V_x = (1/2)[(S_0 + S_{24})V_1 + (S_{12} + S_0)V_2 + (S_{24} + S_{12})V_3] \tag{10}$$

$$V_x = (1/2)[S_0(V_1 + V_2) + S_{12}(V_2 + V_3) + S_{24}(V_3 + V_1)] \tag{10A}$$

$$V_y = (1/2)[(S_{24} + S_{12})V_1 + (S_0 + S_{24})V_2 + (S_{12} + S_0)V_3] \tag{11}$$

$$V_y = (1/2)[S_0(V_2 + V_3) + S_{12}(V_3 + V_1) + S_{24}(V_1 + V_2)] \tag{11A}$$

$$V_z = (1/2)[(S_{12} + S_0)V_1 + (S_{24} + S_{12})V_2 + (S_0 + S_{24})V_3] \tag{12}$$

$$V_z = (1/2)[S_0(V_3 + V_1) + S_{12}(V_1 + V_2) + S_{24}(V_2 + V_3)] \tag{12A}$$

These equations are general expressions for the load voltages where V_1, V_2, and V_3 are the instantaneous voltages at the generator terminals.

If the generator output is a set of balanced sinusoidal voltages

$$V_1 = E \sin \beta \tag{13}$$

$$V_2 = E \sin (\beta - 120°) \tag{14}$$

$$V_3 = E \sin (\beta + 120°) \tag{15}$$

where $\beta = 2\pi f_g t$, f_g is the generator frequency, t is time, and E is the maximum value of generated voltage.

$$V_1 + V_2 + V_3 = 0$$

Therefore, equation 10(A) may be written

$$V_x = -(1/2)[S_0 V_3 + S_{12} V_1 + S_{24} V_2] \tag{16}$$

Substituting equations 13, 14, and 15

$$V_x = -(E/2)[S_0 \sin (\beta + 120°) + S_{12} \sin \beta + S_{24} \sin (\beta - 120°)]$$

But

$$\sin (\beta + 120°) = \frac{\sqrt{3}}{2} \cos \beta - \frac{1}{2} \sin \beta$$

and

$$\sin (\beta - 120°) = -\frac{\sqrt{3}}{2} \cos \beta - \frac{1}{2} \sin \beta$$

Therefore

$$V_x = \frac{E}{2}[(S_{24} + S_0)\frac{\sqrt{3}}{2} \cos \beta + (S_0 + S_{24} - 2S_{12})\frac{1}{2} \sin \beta] \tag{17}$$

The square wave functions S_0, S_{12}, and S_{24} may be expressed in the form of Fourier series[1] as

$$S_0 = (4/\pi)[\sin \alpha + (1/3) \sin 3\alpha + (1/5) \sin 5\alpha + (1/7) \sin 7\alpha + \ldots] \tag{18}$$

$$S_{24} = (4/\pi)[\sin (\alpha - 120°) + (1/3) \sin 3(\alpha - 120°) + (1/5) \sin 5(\alpha - 120°) + (1/7) \sin 7(\alpha - 120°) + \ldots] \tag{19}$$

$$S_{12} = (4/\pi)[\sin (\alpha - 120°) + (1/3) \sin 3\alpha + (1/5) \sin (5\alpha + 120°) + (1/7) \sin (7\alpha - 120°) + \ldots] \tag{19}$$

$$S_{24} = (4/\pi)[\sin (\alpha + 120°) + (1/3) \sin 3(\alpha + 120°) + (1/5) \sin 5(\alpha + 120°) + (1/7) \sin 7(\alpha + 120°) + \ldots] \tag{20}$$

where $\alpha = 2\pi f_s t$, and f_s is the switching frequency.

The functions required in equation 17 are

$$(S_{24} - S_0) = \sqrt{3}(4/\pi)[\sin (\alpha + 150°) + (1/5) \sin (5\alpha - 150°) + (1/7) \sin (7\alpha + 150°) + \ldots] \tag{21}$$

$$(S_0 + S_{24} - 2S_{12}) = -3(4/\pi)[\sin (\alpha - 120°) + (1/5) \sin (5\alpha + 120°) + (1/7) \sin (7\alpha - 120°) \ldots] \tag{22}$$

Substituting equations 21 and 22 into equation 17,

$$V_x = \frac{E\sqrt{3}}{2} \cdot \frac{4}{\pi} \cdot \frac{\sqrt{3}}{2}\Big[\frac{1}{2} \cos (\alpha - \beta + 60°) - \frac{1}{2} \cos (\alpha + \beta + 240°) + \frac{1}{5} \cdot \frac{1}{2} \cos (5\alpha - \beta - 240°) - \frac{1}{5} \cdot \frac{1}{2} \cos (5\alpha + \beta - 60°) + \frac{1}{7} \cdot \frac{1}{2} \cos (7\alpha - \beta + 60°) - \frac{1}{7} \cdot \frac{1}{2} \cos (7\alpha + \beta + 240°) + \ldots\Big] + \frac{E}{2} \cdot \frac{1}{2}(-3)$$

$$\frac{4}{\pi}\Big[\frac{1}{2} \cos (\alpha - \beta - 120°) - \frac{1}{2} \cos (\alpha + \beta - 120°) + \frac{1}{5} \cdot \frac{1}{2} \cos (5\alpha - \beta + 120°) - \frac{1}{5} \cdot \frac{1}{2} \cos (5\alpha + \beta + 120°) + \frac{1}{7} \cdot \frac{1}{2} \cos (7\alpha - \beta - 120°) - \frac{1}{7} \cdot \frac{1}{2} \times \cos (7\alpha + \beta - 120°) + \ldots\Big]$$

$$V_x = (3E/\pi)[\cos (\alpha - \beta + 60°) - (1/5) \cos (5\alpha + \beta - 60°) + (1/7) \cos (7\alpha - \beta + 60°) - \ldots] \tag{23}$$

By a similar derivation it can be shown that

$$V_y = (3E/\pi)[\cos (\alpha - \beta + 180°) - (1/5) \cos (5\alpha + \beta + 180°) + (1/7) \cos (7\alpha - \beta + 180°) - \ldots] \tag{24}$$

and

$$V_z = (3E/\pi)[\cos (\alpha - \beta - 60°) - (1/5) \cos (5\alpha + \beta + 60°) + (1/7) \cos (7\alpha - \beta - 60°) - \ldots] \tag{25}$$

In this analysis the time required to switch from one phase to the next is neglected, for to include it would unnecessarily complicate the expressions. In any final design of a system this switching time must be considered because this represents commutation and causes short interruptions resulting in voltage spikes which must be controlled. However, this is not an extremely difficult problem to overcome.

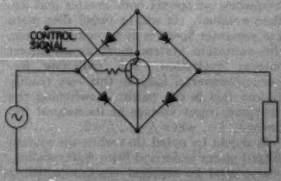


Fig. 10. Bilateral transistor switch

Appendix II. Switching Devices

The power handling capability of the variable-speed constant-frequency system is integrally related to the development of high-power switching devices. Transistors are available in both germanium and silicon materials, but it appears that the inherent higher voltage capability of silicon offers the greater possibility for this application. Silicon transistors are currently available in ratings of 5 amperes peak 300 volts; and a 15-ampere peak 200-volt unit is soon to be released.

A typical circuit using the transistor as a switching device is shown in Fig. 10. While the transistor itself is not a bilateral device, its use in a rectifier bridge as shown makes it essentially so, and as such is capable of switching alternating currents as well as

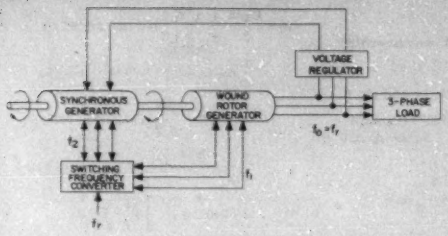


Fig. 11. System diagram of frequency converter used in excitation system for precise frequency system

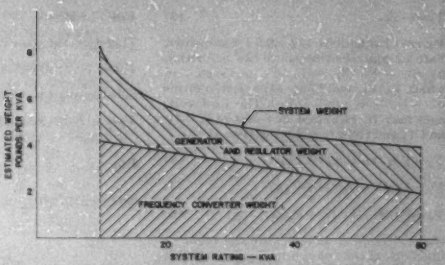


Fig. 12. Estimated weight versus rating of system shown in Fig. 1

direct currents. The switch is turned on and off by means of control current applied to the base of the transistor.

Several manufacturers have made available multiple-junction semiconductor devices[2,3] which exhibit characteristics similar to those of the thyratron. One such device is known as a trinistor. While these devices are not directly interchangeable with transistors, control circuits[2] which permit the use of trinistors as bilateral alternating current switches are available and may be used in the frequency converter. Triuistors in ratings of 300 volts and average current of 15 amperes are currently available, and much higher ratings are soon to be available.

Appendix III. An Additional Electric Power System

One important application of the frequency changer described is in the system shown in Fig. 11 in block diagram form. This system is comprised of a voltage-regulated wound-rotor induction generator receiving its excitation power from a synchronous generator through the static-frequency changer. The wound-rotor machine and the synchronous machine are driven from a common shaft and have the same number of poles. Therefore, if both machines received direct-current excitation they would generate voltages of the same frequency.

Three-phase voltage applied to the rotor of the wound-rotor generator sets up a magnetic field which rotates around the rotor. In this respect the machine acts as a transformer, inducing in the phase windings a voltage proportional to the rotor input voltage. When the rotor of the machine is mechanically driven the angular velocity of the field is changed by an amount equal to the speed of mechanical rotation.

In other words, if the magnetic field rotates at an angular velocity of ω_1 with respect to the rotor, and the rotor is driven at a speed of ω_2, the speed of the field with respect to the stator phase windings will be $\omega_1+\omega_2$. Therefore, the frequency of the output voltage is proportional to $\omega_1+\omega_2$.

The 3-phase voltage applied to the rotor is supplied by the exciter system which comprises the synchronous generator and the switching frequency converter. The synchronous generator, when mechanically driven, produces voltage at a frequency, f_2, proportional to the shaft speed. The output voltage is applied to the switching frequency converter where it is switched at a constant reference frequency, f_r. As previously described, the resulting output voltage from the frequency converter is at a frequency equal to the difference between input frequency and the switching frequency. Therefore

$$f_1=f_r-f_2$$

Since ω_1, the angular velocity of the magnetic field with respect to the rotor, is produced by a voltage at a frequency of f_1, it follows that ω_1 is proportional to f_1, and the frequency of the system output voltage is

$$f_0=f_1+f_2$$

or

$$f_0=(f_r-f_2)+f_2$$

therefore

$$f_0=f_r$$

Thus the system frequency is maintained constant and equal to the reference frequency, f_r, despite changes in shaft speed. With proper control f_r may be maintained constant within limits as precise as one part in 100,000; and, therefore, the output frequency will remain within these same limits.

The ability of the switching frequency converter to put out power of either positive- or negative-sequence at a frequency proportional to the deviation from nominal speed is a very important characteristic of this system. This enables the magnetic field supplied by the excitation system to rotate in either direction with respect to the rotor and thus maintain a constant system frequency, at speeds both greater and less than nominal. It will be noted that when the generator speed is at its nominal value, the excitation system must supply direct current excitation. The switching frequency converter may do this very readily because this is the result of switching the converter input voltage at its natural frequency; i.e., when $f_2=f_r$, $f_1=0$.

It should be noted that while the system output power is derived from shaft rotation, power must be supplied via the switching frequency converter. When the shaft speed is at nominal value, only direct current excitation is required. When the speed deviates from nominal, part of the power is supplied through transformer action in the wound-rotor generator. As the speed deviation increases, the power required from the excitation system increases. The power handling capacity of the switching frequency converter and the synchronous generator, therefore, is a direct function of the speed range.

Nomenclature

E = maximum voltage of generator
f_g = generated voltage frequency
f_0 = system output frequency (Fig. 11)
f_r = reference frequency (Fig. 11)
f_s = switching frequency
f_1 = frequency converter output frequency (Fig. 11)
f_2 = synchronous generator frequency (Fig. 11)
$f_1(t)$, $f_2(t)$, $f_3(t)$ = functions of time defined by Fig. 10(A)
S_0, S_{12}, S_{34} = square wave functions of time defined by Fig. 10(B)
t = time
T_c = conducting time interval of each switch
V_a, V_b, V_c = instantaneous terminal voltages of 3-phase generator
V_x, V_y, V_z = output voltages of switching frequency converter
V_1 through V_6 = instantaneous terminal voltages of 6-phase generator
$\alpha=2\pi f_s t$ = time varying angle of switching operation
$\beta=2\pi f_g t$ = time varying angle of generator voltage
ω_1 = angular velocity of magnetic field with respect to rotor
ω_2 = angular velocity of rotor

References

1. Higher Mathematics for Engineers and Physicists (book), I. S. Sokolnikoff, E. S. Sokolnikoff. McGraw-Hill Book Company, Inc., New York, N. Y., 1941, chap. II.

2. Notes on the Application of the Silicon Controlled Rectifier. *Application Engineering Bulletin*, no. ECG-371-1, General Electric Company, Schenectady, N. Y., Dec. 1958.

3. High Current Trinistors, F. S. Stein, E. W. Torok. *CF58-1397* (available on request).

4. The Rectifier Calculus, W. Morris Goodhue. *AIEE Transactions*, vol. 59, 1940, pp. 587–91.

Discussion

Bert V. Hoard (Boeing Airplane Company, Seattle, Washington): This is a very worthwhile paper which describes another way of obtaining constant frequency from a variable speed generator. However, some additional data on the system, if included in the discussion, will improve its value to the industry. What is the minimum and maximum shaft speed which was used when determining the estimated weight versus rating values shown on Fig. 12 of the paper for the 60-kva generator? What over-all efficiency can be expected for this 60-kva generator at rated load and power factor at minimum and at maximum shaft speed, assuming the generator has the capabilities of meeting present military specifications?

A switching frequency of 2,000 cycles per second rather than 1,200 cycles per second is suggested by the authors as providing higher undesirable frequencies and therefore easier filtering. However, this will increase the number of switching times in each second, and will increase the switching losses and the duty on each semiconductor. Considering these factors and the need for preventing semiconductor failures during abnormal overloads, faults and transient overvoltages, what do they estimate to be necessary derating factors in the semiconductors to provide for reliable operation in the generator?

K. M. Chirgwin and **L. J. Stratton** (Jack & Heintz, Inc., Cleveland, Ohio): The authors are to be congratulated on their semiconductor frequency changer and on the simple mathematical treatment of it that they have developed. In this mathematical treatment the question of commutation has been ignored in order to keep the analysis simple, and this is entirely justifiable. It is unfortunate, however, that no mention has been made of the problems of commutation, since until these problems have been overcome no frequency changer having a substantial power rating can be considered successful. We, therefore, submit the following discussion of the problems of commutation as they appear to exist in this frequency changer.

In order to illustrate the problems of commutation in this frequency changer an analogy with a rotating machine having a commutator will first be developed. The generator frequency-changer combination shown in Fig. 2 of the paper can be represented by a 2-pole direct-current generator having a 6-phase star winding connected to a 6-segment commutator. This machine is rotated at 1,600 revolutions per second. Three brushes (X, Y, Z) spaced 120 degrees apart are provided which feed the power to the 3-phase load. These brushes are not stationary in space but rotate at 400 revolutions per second against or in the same direction as the machine. See Fig. 13. With the directions shown in this figure, the voltage of the brush is a positive maximum when it is at the top of the commutator, is zero when it is on either horizontal center line, and is a negative maximum when it is at the bottom of the commutator.

The mathematical treatment given in the paper requires that the micas and the brushes of this analogy both have negligible width and that commutation be instantaneous. Obviously, in an actual frequency changer the commutation cannot be instantaneous, so let us assume for the moment that the brushes of the analogy have some width. There is now an overlap in time between the currents carried by the incoming and outgoing windings. This is using the "make before break" technique which is used in most commutation problems.

Consider first the conditions that exist when a brush is located on the horizontal center line under the south pole (dotted in Fig. 13). This brush short-circuits segments 5 and 6. The voltages of these segments at the instant shown in Fig. 13 are: $-0.5V_1$ and $+0.5V_1$ respectively; where V_1 is the line to neutral voltage. The voltage that is short-circuited by the brush is the difference of these two or V_1 volts, and is in a direction to cause current to flow out of segment 6 into segment 5. That is in the direction required to transfer the out flowing load current from segment 5 to segment 6, (provided the speed of rotation of the brush is less than that of the commutator). As in any commutation process, this transfor of current will be opposed by the leakage inductance of both of the windings so that the transfer will take a certain time for completion. If the width of the brush is correctly chosen, the transfer will be just complete when segment 5 breaks contact with the brush, and perfect commutation results.

Since the time required for commutation varies directly with the load current but the voltage causing commutation remains fixed, it is necessary to vary the width of the brush (overlap time) with the load current to ensure perfect commutation.

In addition, the magnitude of the voltage between adjacent commutator segments which is the voltage causing commutation varies (sinusoidally) with brush position being a maximum on the horizontal line and zero at the top and bottom of the commutator, so it is also necessary to vary the overlap time with brush position in order to achieve perfect commutation.

If steps were taken to ensure the correct variation of overlap time with load current and brush position, then perfect commutation would be possible for an outflowing load current and for all positions of the brush under the south pole. When we consider conditions under the north pole, however, we find the direction of the voltage that is short-circuited by the brush is such as to oppose the transfer of current, and commutation cannot be accomplished without sparking. If the direction of rotation of the brushes is opposite to that of the commutator then the situation is unchanged because the direction of motion of the brushes on the commutator is unchanged.

Returning now to the semiconductor frequency-changer, it appears that by using the "make before break" technique satisfactory commutation could be achieved for half the time by varying the overlap in an appropriate manner, but for the other half of the time some external means of forcing the commutation would be necessary.

Another way of dealing with the situation that has been used where small powers are involved is to employ a "break before make" switching sequence. With this approach the energy stored in the inductance of the coil is dissipated (in an arc for example) each time a switch is opened. If the switch used is the bilateral transistor switch shown in Fig. 10 of the paper, then the energy stored in the inductance of the coil appears as heat in the transistor. When the power rating of the frequency changer becomes large, this extra heat in the transistor is objectionable because of the reduction in efficiency and because of the de-rating of the transistor. We find with a normal design of the generator this extra loss can be equal to the normal transistor losses, which means the power rating of the transistor is reduced to about one half normal.

We would be interested to know if a "break before make" technique is used in this frequency changer and if so how much the transistors must be de-rated; or whether some external means of commutation are employed.

It should be pointed out that the commutation difficulties described do not occur with all semiconductor frequency changers. A frequency changer using silicon controlled rectifiers as the switching device has been developed by the discussers.

This frequency changer is described in a paper entitled "Precise Frequency Power Generation from an Unregulated Shaft," by K. M. Chirgwin and L. J. Stratton, which is scheduled for presentation at a forthcoming AIEE meeting. Also, a paper[1] describing the application of the frequency changer to an aircraft variable-speed constant-frequency generating system was published recently.

The silicon controlled rectifier used in this frequency changer cannot be turned off by the control circuit. Because of this, commutation is very similar to that occurring in a normal (diode) rectifier circuit, where the firing of the incoming controlled

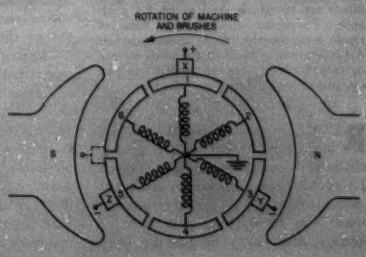


Fig. 13. Rotating machine analogy of the frequency changer and generator

rectifier causes the load current to transfer from the outgoing to the incoming controlled rectifier. The length of the overlap period automatically adjusts to the time required to complete the transfer.

The frequency changer has further advantages in that controlled rectifiers presently are available in higher power ratings than transistors (they may always be so[3,4]), and in that it has the capability of voltage regulation within itself; a characteristic that seems to be absent in the authors' frequency changer.

REFERENCES

1. VARIABLE-SPEED CONSTANT-FREQUENCY GENERATOR SYSTEM FOR AIRCRAFT, K. M. Chirgwin, L. J. Stratton. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, Nov. 1959, pp. 304–10.

2. PNPN SWITCHES, J. M. Goldey. *CP58-1359* (not available).

3. See reference 3 of the paper.

4. VARIABLE-SPEED SYNCHRONOUS-FLUX GENERATOR, V. DeBiasi. *Space/Aeronautics*, New York, N. Y., Apr. 1959.

R. D. Jessee and William J. Spaven: The commutation problem pointed out by Messrs. Chirgwin and Stratton is one of which we are aware. Semiconductor switches have the characteristics of faster turn-on than turn-off. Therefore, the normal mode of operation of the frequency converter described allows some overlap in switching; i.e., one set of switches turns on before the previous set turns off. This type of operation causes a short circuit on the generator through the switches each time the switches operate. As pointed out, this type of operation requires some derating of the semiconductor switches. As a set of switches completes the turn-off operation, the short circuit is removed and the result is a transient voltage of very short duration; a spike. By delaying the turn-on, the overlap may be eliminated and therefore short-circuiting of the generator does not occur. Instead, the generator terminals are opened momentarily, interrupting the current flow. This, of course, causes voltage spikes but without the short-circuiting of the generator through the switches.

In order that the switches may operate efficiently, the voltage spikes must not appear on the switches. The voltage spikes may be suppressed by providing a path for the current between the generator and the switches. By use of energy storage circuits, we have been successful in limiting the spikes appearing at the switches on models built and tested. Apparently very little de-rating of semiconductors is required because of commutation.

The questions raised by Mr. Hoard pertain to the design of specific apparatus. While the intent of the paper was to present the principles of a method for accomplishing frequency conversion, the questions raised are pertinent to the application of the principles described. Fig. 12 was intended to give a rough estimate of the weight of a constant-frequency electric power system having a two-to-one speed range. The estimate was based on experience in electric conversion equipment and aircraft generator manufacture.

Because the generator currents are continually switched, the root-mean-square values of these currents are increased for a given power output. For this reason, the generator efficiency can be expected to be somewhat lower than for a constant-speed synchronous generator. Further, efficiency depends on the specific design employed in the generator and switching circuits. Calculations indicate a generator efficiency in the order of 80% in this system.

In the final design of a system using a frequency converter as described here, the actual generated and switching frequencies are subject to several compromises. There are limitations to not only the switching frequency but also generator speed and, therefore, generator frequency. More phases may be used in generation, and by so doing, the lowest frequency of undesirable voltage components may be increased in proportion. This will result in a compromise where the generator and switching frequencies may be lowered and at the same time gain high-frequency undesirable voltage components with reduced amplitude. This type of design could work in favor of both generator and frequency converter. Of course, any such compromise must consider the number of switches which must be added in order to accomplish the final configuration. In considering derating for semiconductors, all operating conditions must be considered. The power loss caused by switching will be approximately the same for each switching operation at any particular load. Therefore, the switching losses will vary approximately with the switching rate.

Characteristics and Measurement of Ripple in Aircraft Electric Power Systems

OSCAR MARKOWITZ
MEMBER AIEE

Synopsis: This paper describes the work accomplished at NADEVCEN (the Naval Air Development Center) to develop definition of frequency characteristic of ripple and to provide guidance in the voltage measurement of ripple. Factors influencing the unpredictable variations in ripple measurements are recognized and noted for guidance. Frequency characteristic data of ripple accumulated from various aircraft d-c power sources formed the basis for the developed limits recommended to be integrated into military specifications.

RIPPLE in aircraft electric systems, like other aircraft electric system characteristics, has been taken for granted. Its existence has been known and limits delineated in various military specifications[1–4] as a maximum peak from an average assumed to be the d-c level. Tremendous increases in the sophistication and performance of military electronic equipment has forced the need to know more about the characteristics of ripple. This paper deals primarily with various aspects of ripple, its measurement and development of definition for the frequency characteristics of the ripple.

The peak value of ripple as a characteristic is not as valuable to equipment designers as the voltage at any given frequency within the frequency spectrum of that ripple. Previous efforts to determine this frequency spectrum characteristic of power system ripple have been fragmentary.[5] In a sincere effort to determine susceptibility to ripple, some equipment designers have superimposed single-frequency a-c components on the d-c input to their equipment. By varying the a-c component in amplitude and frequency, they have evaluated the sensitivity of their equipment to the a-c component. The missing link in this type of evaluation was a knowledge of the limits of frequency characteristic of ripple obtained in power systems. Assumptions have been made in the past that a specified peak value of ripple could be coincident of a same maximum ampli-

Paper 59-879, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 24, 1959; made available for printing April 22, 1959.

OSCAR MARKOWITZ is with the United States Naval Air Development Center, Johnsville, Pa.

The laboratory work in obtaining measurements and pictures of ripple was accomplished by Howard Ireland. The laboratory work in obtaining data for the frequency characteristics of ripple was accomplished by Henry Kunicki. The author gratefully acknowledges the information he received on ripple of various equipments from the Naval Air Center at Patuxent River, Md. The author also wishes to thank the many engineering groups who have been extremely co-operative in establishing limits for aircraft electric system characteristics and the author's investigation of electric systems-utilization equipment compatibility problems. Illustrations are official photographs of the United States Navy.

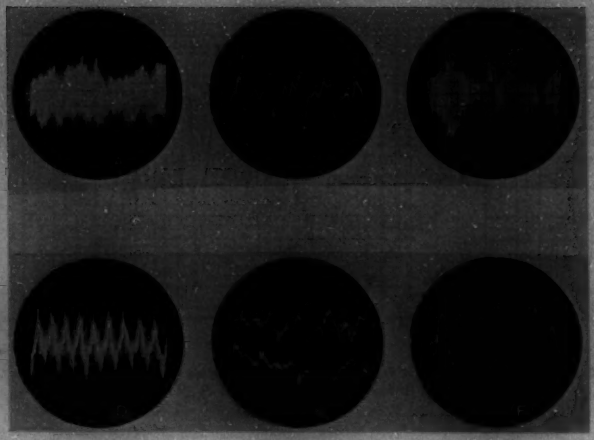


Fig. 1. Sample oscillograms of d-c generator ripple. Generator: Manufacturer D, 200 amperes, 4,000 to 8,000 rpm

A—4,000 rpm, no load, 360 mechanical degrees of generator rotation, 2 milliseconds per division, 0.5 volt per division

B—4,000 rpm, no load, bar to bar, 100 microseconds per division, 0.5 volt per division

C—8,000 rpm, no load, 360 mechanical degrees of generator rotation, 750 microseconds per division, 0.5 volt per division

D—4,000 rpm, full load, 360 mechanical degrees of generator rotation, 2 milliseconds per division, 0.2 volt per division

E—4,000 rpm, full load, bar to bar, 100 microseconds per division, 0.2 volt per division

F—6,000 rpm, no load, bar to bar, 50 microseconds per division, 0.5 volt per division

Table I. Maximum Peak Ripple Obtained From Various Equipments

Manufacturer	Equipment	Ampere Rating	Volts Ripple
G	Unregulated TR	200	2.0
G	Unregulated TR	100	1.3
G	Unregulated TR	50	4.5
G	Unregulated TR	25	5.8
B	Unregulated TR	100	1.1
B	Unregulated TR	50	1.15
A	Unregulated TR	100	1.3
A	Unregulated TR	50	1.0
C	Unregulated TR	100	1.23
J	Unregulated TR	200	2.45
B	Regulated TR	50	1.8
E*	Regulated TR	50	0.6
H	Regulated TR	200	2.4
I	Regulated TR	50	2.45
B	Generator	200	1.53
B	Generator	300	0.69
D	Generator	200	1.4
D	Generator	400	1.33
F	Generator	400	0.85

* Environmental evaluation not completed.

Note: Volts ripple is the maximum peak ripple voltage ± from the d-c average measured for all conditions of operations.

tude at any frequency from close to direct current to well above the audio spectrum. This assumption has forced severe filter penalties on some equipment. NADEVCEN made an investigation to provide a basis of delineating the frequency characteristic of ripple.

Influencing Factors of Ripple

In present aircraft and missiles d-c power is obtained directly from d-c generators and batteries and indirectly through transformer rectifiers from a-c power systems. This paper considers the ripple in 28-volt d-c power systems, including d-c generators and transformer rectifiers.

Ripple in d-c generators is primarily due to commutation at the armature. The variation of flux across the pole faces, pole to pole, and of winding to winding give rise to the lower frequency components. Arcing at the commutator gives rise to the higher frequency components. The arcing is function of design (effectiveness of flux shift compensation as a function of load), load, service condition of commutator and brushes, and filter. Input speed variations can shift the lower frequency components but has relatively little effect on the higher frequency components due to arcing. A battery placed across the generator is very effective in reducing the peaks of ripple above the average d-c level.

The fundamental aspect of ripple in transformer rectifiers (TR) is related to flow of current only occurring during the peak of each sine wave. The conduction angle for each phase of a 6-phase half-wave TR is 60 degrees, 30 degrees each side of the sine-wave peak. Practical considerations result in distortion of each 60-cycle sine-wave peak and in variations in d-c component from one sine wave peak to the next through the six making up 360 degrees of the fundamental frequency. Thus the TR ripple is affected from its theoretical limits by the variation between rectifiers, phase-to-phase transformation, and windings. Extreme emphasis placed on minimum weight for aircraft TR's has further emphasized internal phase to phase variations and no-load to full-load variations. These factors become more acute for regulated transformer rectifiers since regulation is usually obtained by varying the conduction angle for each rectifier from 60 degrees to some value less than 60 degrees. The a-c power source can affect ripple with unbalanced voltages between phases and when the source impedance is relatively high in relation to the TR conduction angle load currents. The power source can further influence the ripple when nonlinear loads on the same source influence a-c waveform in the region of the TR conduction angle. It has been recognized that a TR in itself is a nonlinear load on the a-c power system and does increase the harmonic content of the a-c power. With the narrowing of the conduction

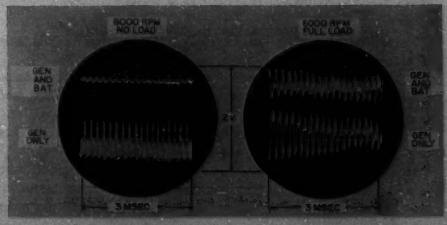


Fig. 2. Influence of a battery on d-c generator ripple. Generator: Manufacturer B, 200 amperes d-c

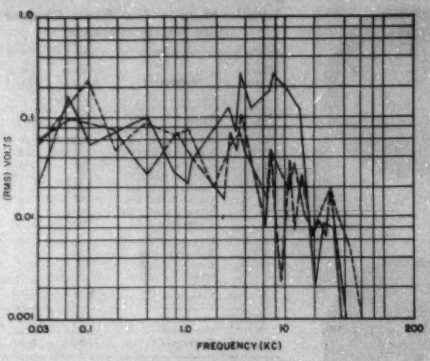


Fig. 3. Frequency characteristics of ripple from a 300-ampere generator. Manufacturer A

——— No load
— — Half load
- - - Full load

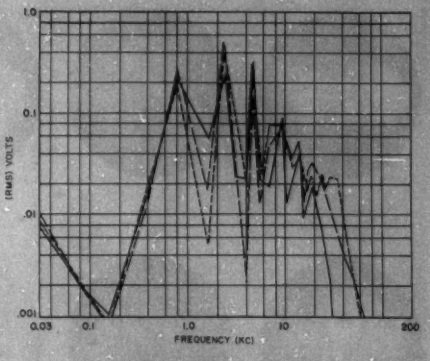


Fig. 5. Frequency characteristics of ripple from a 100-ampere 6-phase unregulated TR. Manufacturer A, 3-volt input balanced

——— No load
— — Half load
- - - Full load

angle in regulated units, the influence on harmonics becomes progressively worse. Increased a-c source impedance for a given TR load also correspondingly increases influence on harmonics. Voltage modulation,[6,7] which can ride through a TR and become part of the ripple is also found in a-c systems. Thus minimum ripple from a TR can be achieved only by both good design with close production control and careful application considerations.

Measurement of Ripple

Measurement of ripple in d-c power systems for aircraft has been in the past predicated on the need to know how far the ripple caused voltage to leave the average d-c level, and is referred to in the aircraft power field as a maximum ripple voltage. It has been specified and measured in military specifications[1-4] as a maximum peak voltage. This is contrasted to other fields where ripple is considered in terms of per cent. The per cent figure is confused, since it is obtained for other power fields[8] by comparing the rms value of ripple to average d-c value and for electronic fields[9] as the total of plus and minus peak deviations compared to the average d-c value. The various methods of measurement cannot be correlated for aircraft d-c systems, since the ripple is almost always a very complex wave. Other factors that have entered into the confusion of ripple measurement are as follows:

1. Peak-to-peak voltmeters have been confused with peak reading voltmeters. The peak reading voltmeter reads only the voltage from an a-c zero to the peak going in the positive direction for one polarity connection and from a-c zero to the peak going in the negative direction for the opposite polarity connection. The a-c zero may or may not coincide with the average direct voltage. The peak-to-peak reading voltmeter measures the span in voltage between the negative ripple peak value and the positive ripple peak value.

2. Time constants of peak reading or peak-to-peak instruments vary over considerable limits. Short-duration peaks with random or relative long duration of re-occurrence is often obtained in generator ripple. In many instruments, unless the voltage to be measured reoccurs on a cyclic basis of high enough repetition rate, the readings do not reflect the maximum peaks.

3. It is not always clear when using a vacuum-tube voltmeter that the actual reading obtained is a function of average, rms, or peak of the input wave. This confusion is compounded when the meter scale is marked as a reading not corresponding to the function of measurement. On a sine wave, if the function of measurement is known, conversion can be obtained. With a complex wave, if the meter is not calibrated in terms of the function of measurement, reading cannot always be corrected by conversion factors.

4. When using a peak reading meter, care should be exercised that the accuracy is valid for the low voltages normally encountered in ripple, 0.1 to 5 volts. Many instruments use diodes in their input. When input voltages are measured which are in the same order of magnitude as the conduction knee of the diode, then the reading lies somewhere between the peak and average values. This uncertainty can be eliminated if the ripple is amplified to the order of 100 volts and the peak reading divided by the amplification factor.

In order to obtain repeatable results in voltage measurements of TR ripple, the effect of a-c source upon ripple, and the

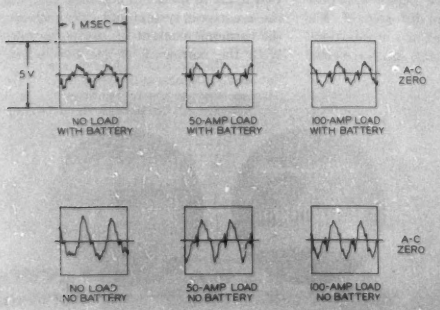
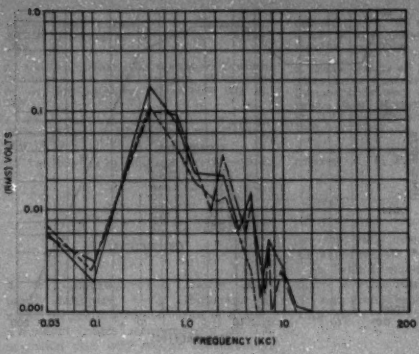


Fig. 4. Influence of a battery on TR ripple. Battery: 24 volts, 36 ampere-hours. TR: Manufacturer A, 100 ampere, unregulated 6-phase



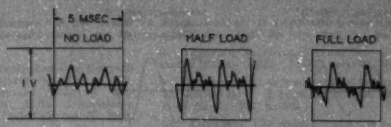


Fig. 6 (left). Frequency characteristic of ripple from a 30-ampere 6-phase regulated TR. Manufacturer E

No load
Half load
Full load

Fig. 7 (above). Ripple from a 30-ampere 6-phase regulated TR. Manufacturer E: Ripple filter choke input with 1,500-microfarad capacity

effect of the TR upon a-c source, waveform should be recognized. To allow consistent ripple voltage measurements, at least 360 mechanical degrees of generator rotation and at least 0.1 second of TR operation should be considered in the measurement.

Frequency characteristics of ripple have been measured at NADEVCEN using conducted radio-interference techniques. Military specification[10] for such measurements start at 140 kc and go up in frequency. Frequency characteristics of ripple picks up where conducted radio interference stops (140 kc) and goes down in frequency to 30 cps (cycles per second). The lower 30-cps limit is based on the general lack of frequency components below 30 cps, the limitation of available instruments to measure conducted interference below 30 cps, and the limits of frequency characteristics of voltage modulation established for a-c systems.[6]

A screen room was used for all frequency characteristics measurements. Generators are driven with the generator mounting pad face coincident with the wall of the screen room. This allows the heavy drive to remain outside the screen room. The mechanical arrangements of loading and power leads were consistent with established conducted radio interference measurements techniques. Due to the low frequencies measured and extremely low impedance sources, standardizing impedance (stabilizing) network could not be used. Measurements in the range of 30 cps to 15 kc were made with a *NM-40A* Radio Interference Field Intensity Meter. Measurements in the range of 14 kc to 140 kc were made with a *URM-6* Radio Interference Field Intensity Meter. Both instruments are manufactured by Stoddart Aircraft Radio Company. As a check on the measurements, for each frequency that was measured a sine wave of corresponding frequency was substituted from an oscillator. The sine wave was varied until its output provided a reading identical to the ripple measurement. The voltage of this sine wave was read independently and compared with the voltage read for ripple in the test set.

Table II. Source Influence Upon Ripple of Manufacturer "G" 100 Amp Unregulated Transformer-Rectifier

Input	Additional A-C System Load	TR Load	Ripple Plus	Ripple Minus
30 kva a-c generator, *B* manufacturer	None	0	0.40	0.20
	None	10	1.10	1.50
	None	60	1.25	1.10
	None	100	1.35	0.90
15 kva a-c generator, *B* manufacturer	None	0	0.33	0.25
	None	10	1.90	2.10
	None	50	1.80	1.50
	None	100	1.80	1.90
	Manufacturer *A* 200-ampere unregulated TR with 50-ampere load	0	0.35	0.20
		10	1.65	1.85
		50	1.30	1.35
		100	1.50	1.50
	Manufacturer *I*, 50-ampere regulated TR with 50-ampere load	0	0.30	0.20
		10	2.90	5.70
		50	2.10	5.40
		100	1.90	4.80

A-c input maintained at 118 volts 400 cps.

Notes: Data for this table were obtained from the Naval Air Test Center, Patuxent River, Md. Ripple is peak voltage measured from d-c average.

Test Data and Discussions

Fig. 1 shows oscillograms of ripple from a d-c generator. These oscillograms are fairly characteristic of ripple obtained from aircraft d-c generators. There is some variation in ripple with different input speeds. Variation in ripple due to load change is very pronounced in both peak values and frequency content.

An aircraft battery placed across a d-c generator lowers the ripple to quite a degree. The battery is not quite as effective for the full-load operation as for the light-load operation. The oscillogram shown in Fig. 2 illustrates the effect of a battery upon ripple.

The military specification for d-c generators[1] limits peak ripple voltage to 1.5 volts. The co-ordinating military specification for regulators[2] permits the generator peak ripple voltage to increase to 2.1 volts due to the regulator. Tests made in NADEVCEN laboratory failed to reveal any significant change in ripple due to the addition of a regulator. The peak ripple data shown in Table I for representative d-c generators were obtained using carbon pile regulators qualified to *MIL-R-6809*.[3]

Table III. 30-Ampere Regulated Transformer-Rectifier Influence on Input A-C Waveform

Harmonic	30-Kva Generator: Manufacturer B, Per Cent Harmonic With No Load on Generator	30-Kva Generator: Per Cent Harmonic With Fully Loaded Manufacturer B TR on Generator	15-Kva Generator: Manufacturer B, Per Cent Harmonic With No Load on Generator	15-Kva Generator: Per Cent Harmonic With Fully Loaded Manufacturer E TR on Generator
2		0.71		0.20
3	0.25	0.66	0.12	1.34
4		0.66		0.60
5	0.6	5.20	1.85	2.90
6				0.33
7		1.55	1.65	1.34
8		0.89		0.52
9	0.2	0.60		1.34
10				0.64
11		2.20	0.39	1.00
12				0.63
13				1.26
14				0.43
15	0.26	0.63		1.00
16				0.62
17		1.50		0.55
19				0.52
20		0.80		
23		0.97	0.6	0.6
30		0.71		

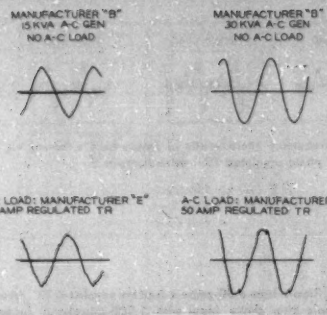


Fig. 8. Regulated TR influence on input a-c waveform

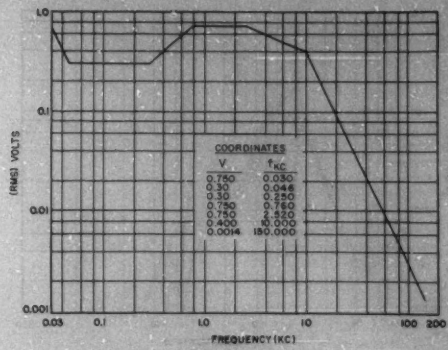


Fig. 9. Recommended limits for the frequency characteristics of ripple

A representative frequency characteristic of ripple for a d-c generator is shown in Fig. 3. It should be noted that all the frequency characteristic curves are representative of the maximum readings during one set of operating conditions only at the corresponding frequencies and do not show the readings obtained at frequencies other than the maximum points. Most of the readings obtained for frequencies other than the maximum points were below the 0.001-volt level.

Fig. 4 is illustrative of ripple obtained from an unregulated TR. The addition of a battery across the TR has little effect except at the no-load operation. The effect is much less than when a battery is placed across a generator. However, for TR's with higher ripple voltage the battery has a corresponding increasing influence in reducing ripple. Table I lists peak readings of ripple obtained from various TR units. These readings for TR's are the maximum obtained with different environments and loads.

The effect of unbalance in the a-c input to a TR on the frequency characteristic is shown in Fig. 5. The 800-cycle component due to the 3-volt input unbalance has increased from a negligible value for balanced input to one of the largest components. The unbalance was a 3-volt spread between the highest and lowest line to neutral voltages. The same TR with no input unbalance showed an almost identical frequency characteristic with the exception of the 800-cps component. Regulated TR's have inherently high output ripple in some cases measuring as high as 18 volts. Use of a filter becomes a necessity to reduce the ripple to acceptable limits. The frequency characteristic of one such well filtered regulated TR is shown in Fig. 6. An oscillogram of ripple from the same unit is shown in Fig. 7.

The TR, when used in air-borne electric systems, is usually a reasonable percentage of load on the a-c generator. Being a heavy nonlinear load there is an effect on a-c waveform. The waveform distortion in turn affects ripple. Table II shows the different ripple reading obtained with a given TR and different generators. This table illustrates how the addition of other nonlinear loads on the a-c generator further increases the ripple. Fig. 8 shows the effect on a-c source waveforms of regulated TR's. Table III shows the harmonic analysis of the waveforms shown in Fig. 8. The two regulated TR's had different methods of obtaining regulation. Manufacturer *B* provided the control in the input transformer while manufacturer *E* provided control by switching impedances in series with the output of the transformer. The effect upon a-c waveform was much greater when the control was integral with the input transformer.

Conclusions and Recommendations

From independent studies it has been concluded that many air-borne utilization equipment are at the threshold of ripple tolerance when the peak ripple voltage goes above 1.5 volts. Most utilization equipments desire peak values of ripple below the 1.5-volt limit. However, present state of the art in a-c power equipment with reasonable care in design, quality control, and proper application can only confine ripple to the maximum limit of 1.5 volts. Measurement of the peak ripple voltage must be made with extreme caution. It is usually best to monitor and check the measurement with an appropriately calibrated oscilloscope capable of appropriate sweeps and having a minimum frequency response between 10 cps and 100 kc.

The study made and the data obtained at NADEVCEN indicate that realistic limits shown in Fig. 9 can be established for the frequency characteristic of ripple. These limits bridge the gap between the lower end of conducted radio interference and the upper end of frequency characteristics of a-c modulation.

Ripple measured during laboratory testing cannot be considered as an absolute value but only an indication to assist in the final application of the TR to an air-borne electric system. The final application has to consider both the influence of TR upon a-c system waveform and the influence of the a-c system waveform upon ripple.

The following recommendations are made:

1. The maximum peak ripple voltage be established as 1.5 volts maximum deviation from the d-c average voltage.
2. Measurements of ripple be made with due cautions presented in this paper and be monitored with an appropriate oscilloscope.
3. The frequency characteristic of ripple shown in Fig. 9 be established as maximum limits and be integrated into appropriate specifications.
4. Measurement of ripple and application of TR's be made with due cognizance of the TR's influence upon a-c system waveform and the a-c system waveform's influence upon ripple.

References

1. GENERAL SPECIFICATION FOR AIRCRAFT CONVERTERS. *Military Specification* no. *MIL-C-7115B*, Bureau of Aeronautics, Washington, D. C., Jan. 2, 1957.

2. GENERAL SPECIFICATION FOR AIRCRAFT ENGINE

Driven Direct Current 30 Volt Generator. *Military Specification no. MIL-G-6162-A1*, Sept. 9, 1955.

3. General Specification for Direct Current Generator 30 Volt Voltage Regulator. *Military Specification no. MIL-R-6809*, July 14, 1950.

4. Characteristics of Aircraft Electric Power. *Military Specification no. MIL-E-7894A*, May 17, 1955.

5. The Nature of Voltage Ripple on D-C Generators. F. N. Collamore. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, Jan. 1955, pp. 390–93.

6. Development of Definitions for Voltage Modulation Characteristics in 380- to 420-Cycles-Per-Second Aircraft Electric Systems. Oscar Markowitz. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 75, Sept. 1956, pp. 193–97.

7. Voltage Modulation on Aircraft Power Systems. R. E. Kiskow, C. F. Yoke. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, Nov. 1957, pp. 314–19.

8. American Standard Definitions of Electrical Terms (book). AIEE sect. 65.11.260.

9. Fundamentals of Engineering Electronics. William G. Dow. John Wiley & Sons, Inc., New York, N. Y., 1937, p. 508.

10. Aircraft Electrical and Electronic Equipment Interference Limits and Tests. *Military Specification No. MIL-I-6181B*, May 29, 1953.

Discussion

E. W. Colehower (The Martin Company, Baltimore, Md.): Paragraph 2, under the heading Measurement of Ripple: Consistent ripple measurements of TR operation need only require measurement over a time interval of 1 cycle rather than 0.1 second. Generator ripple measurements, as proposed, are limited to 360 mechanical degrees, which would be 0.02 second for a generator at 3,000 rpm. A 60-cps TR measurement should be for 0.017 second and a 3,200-cps TR measurement for only 0.31 millisecond. These are comparable intervals and should include all nonrepeatable characteristics.

In Table I, manufacturer *A*, 100-ampere rating 1.3-volt ripple (assumed maximum deviant from average d-c) does not agree with Fig. 4 (2.0 volts).

Oscar Markowitz: The measurement of ripple must include any subharmonics and the voltage modulation components which are inherent with a-c power sources. Examination of Figs. 5 and 6 shows that the frequency components below the 1-cycle interval start to increase as frequency goes to 100 cps. Reference 6 indicates that the low-frequency components due to modulation may reach a maximum about 10 cps. Thus the interval of ripple measurement should observe a 10-cps component (0.1 second).

Fig. 4 is an artist's sketch of the actual oscillogram. Measurements made from the oscillogram show agreement within instrumentation error (±5%). This agreement also considers that the data were obtained from different units. The data of Table I were obtained from a unit undergoing qualification tests at one government activity. The oscillogram for Fig. 4 was obtained from a different unit at a second government activity.

Nuclear Radiation and Electronic Equipment

J. R. CRITTENDEN
ASSOCIATE MEMBER AIEE

NUCLEAR radiation causes material degradation through several processes. Ionizing radiations, the charged particles, and gamma rays cause organic materials to change chemically and electrical conductivity to increase. Energetic neutrons displace atoms by knocking them out of their position in the structure of the material; and the displaced atoms, as charged particles, also cause ionization and secondary displacements. Less energetic neutrons may be absorbed by a nucleus. The new isotope may be unstable and decay in any of several processes such as fission, alpha, beta, or gamma emission, each of which may cause ionization. Material degradation then is caused by ionization, displacement, and transmutation which are observed as chemical changes and increased conductivity.

Increased conductivity is a function of the current carriers available and their ability to move. (Mobility is a characteristic of the material under consideration.) Ionization generates large numbers of current carriers (electrons and ions or holes). The degree of ionization is a direct function of the intensity of the nuclear flux. The increase in electrical conductivity is thus related to the intensity of the nuclear flux as illustrated in Fig. 1.[1]

Chemical changes, on the other hand, are a function of the time integral of the dose-rate. Ionization, displacement, and transmutation are discrete events. For a given flux intensity (dose-rate) the number of events per unit time can be calculated. Because the reaction during an event can "go" in any of several processes, it is difficult to define the number

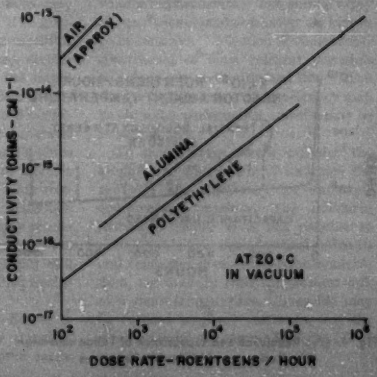


Fig. 1. Variation of conductivity as a function of dose rate

Paper 59-873, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 23, 1959; made available for printing May 28, 1959.

J. R. Crittenden is with General Electric Company, Owensboro, Ky.

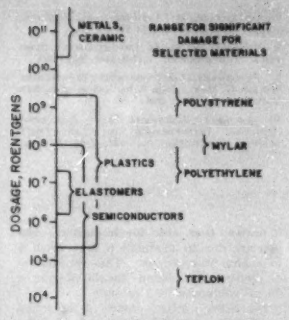


Fig. 2 (left). Scale of relative tolerance of materials

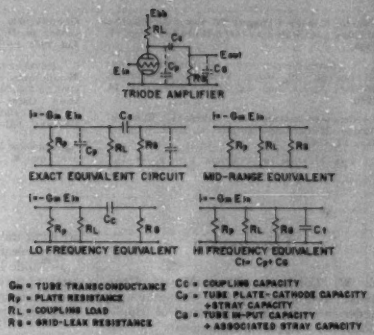

TRIODE AMPLIFIER
EXACT EQUIVALENT CIRCUIT
MID-RANGE EQUIVALENT
LO FREQUENCY EQUIVALENT
HI FREQUENCY EQUIVALENT
Gm = TUBE TRANSCONDUCTANCE
Rp = PLATE RESISTANCE
RL = COUPLING LOAD
Rg = GRID-LEAK RESISTANCE
Cc = COUPLING CAPACITY
Cp = TUBE PLATE-CATHODE CAPACITY + STRAY CAPACITY
Cg = TUBE IN-PUT CAPACITY + ASSOCIATED STRAY CAPACITY

Fig. 3 (right). Electronic amplifier and equivalent circuits

of events required to bring about a specific effect, but a range of the time integrals of dose-rate (dose) can be determined. Fig. 2 illustrates the range of doses to obtain an effect for several materials.[2] As only a small portion of the induced changes revert, chemical changes are permanent in contrast to the transient conductivity changes.

Chemical changes and increased conductivity in electronic equipment can result in failure or malfunction. A simple audio amplifier may be used to demonstrate radiation effects in electronic component. Fig. 3 shows an amplifier and its equialent circuits. As indicated by the equivalent circuit, the transfer function of the amplifier, the ratio of the output signal to the input signal, E_{out}/E_{in}, is determined by the operating characteristics of the vacuum tube and its associated resistors and capacitors. The use of equivalent circuitry implies that the characteristics of the components remain constant. In practice, these values may vary; but the designer, through experi-

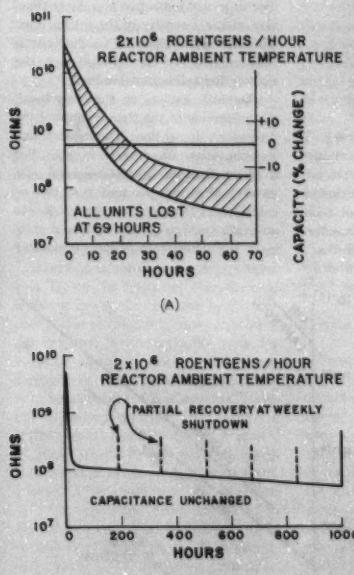
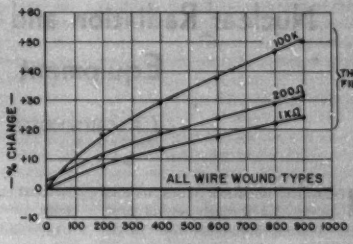
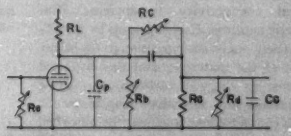


Fig. 4. (A) Resistance and capacitance of Teflon capacitors. (B) Resistance of Mylar-film capacitors

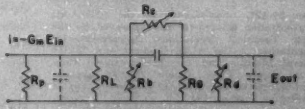
Fig. 5 (right). Typical behavior of resistors irradiated in the Oak Ridge National Laboratory (ORNL) graphite reactor at reactor ambient temperature. Total integrated dose: 2×10^9 roentgens





Gm = TUBE TRANSCONDUCTANCE
Rp = PLATE RESISTANCE
RL = COUPLING LOAD
Rg = GRID-LEAK RESISTANCE
Cc = COUPLING CAPACITY
Cp = TUBE PLATE-CATHODE & STRAY CAPACITY
Cg = TUBE IN-PUT CAPACITY + ASSOCIATED STRAY CAPACITY
Ri = INPUT STRAY LEAKAGE
Rb = PLATE-CATHODE STRAY LEAKAGE
Rc = CAPACITOR SHUNT LEAKAGE
Rd = OUTPUT STRAY LEAKAGE

Fig. 6 (right). Electronic amplifier showing radiation-induced current paths



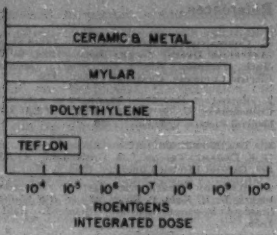


Fig. 7 (left). Life expectancy of amplifier containing various materials

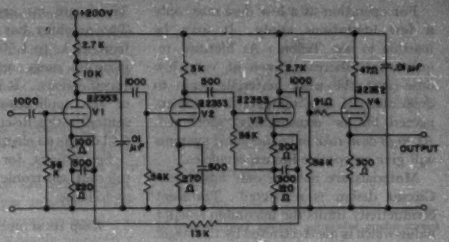


Fig. 8 (right). Radiation tolerant amplifier

ence, has learned to accommodate the inconsistencies. The problem is to evaluate the radiation-induced changes and learn to accommodate them.

The amplifier depicted in Fig. 3 has three elements which will be considered: the capacitor, the resistor, and the vacuum tube. A number of studies on the tolerances of vacuum tubes to nuclear radiation have been reported.[2-5] Several types were found which are essentially unaffected by radiation to doses which exceed the life expectancy of the other amplifier components. Wire wound resistors are also insensitive to nuclear radiation. Large value capacitors are apparently the most sensitive component.

Fig. 4 shows the characteristic behavior of two commercial capacitor types. The behavior of capacitors having Mylar and Teflon dielectrics illustrate the effect of nuclear radiation; that is, large changes of insulation resistance, dielectric conductivity, and little or no change of capacitance. As noted in Fig. 2, Teflon is quite susceptible to radiation damage whereas Mylar is not so easily affected. Both capacitor types were exposed to the same environment and in sufficient quantity to achieve reliable results.

The behavior of resistors is shown in Fig. 5. Wire wound devices are not seriously affected, but in the thin-film varieties the effect of displacements is to distort the film causing a reduction in conductor cross-sectional area. As illustrated in the figure, resistance values increased. It is advantageous to use wire wound devices when the required values and sizes are available. (Because they are temperature sensitive and vary with age, carbon composition resistors are not considered for this application.)

The foregoing data indicate the effects induced by nuclear radiation on the components of the amplifier circuit. Fig. 6 shows the initial circuit modified by the variables caused by nuclear radiation and the modified equivalent circuit.

It should be noted that catastrophic failure of the amplifier occurs when a capacitor develops a short circuit. The same rule is true for the tube or the resistors; that is, drastic nonreversible changes are considered component failure and will cause amplifier failure. As they may be accommodated in the circuit design, repeatable or reversible changes do not necessarily constitute component failure.

Analysis of the equivalent circuit in Fig. 6 indicates that the radiation-induced current paths do not appreciably alter the amplifier's transfer function except at low frequencies. Should the input resistor paralleling R_a have a high ohmic value, it is possible to obtain some attenuation of the input signal; or should resistor R_g of the amplifier have a high value, the parallel stray leakage may cause some attenuation. The more important leakage path shunts the coupling capacitor. The coupling capacitor is often called the blocking capacitor as one of its purposes is to restrict direct-current flow. When the blocking capacitor conducts, the d-c voltage level at the amplifier output is increased. Normal amplifier design assumes that no direct current flows through the coupling capacitor. R_g values are usually greater than 10^6 ohms. It can be seen from the data in Fig. 4 that capacitor shunt resistance may decrease to 10^7 ohms or less. The grid-leak resistor and the blocking-capacitor leakage resistance form a voltage divider which applies a d-c voltage to the amplifier output or to the next stage of the circuit. The magnitude of the d-c voltage is a function of the resistance ratio R_c/R_g. As R_c may be as low as 10^6 ohms and R_g is frequently 10^8 ohms or more, it is possible to apply 50% of the plate voltage at the amplifier output. Although the a-c fed into the amplifier may not be seriously distorted by its action, the variable d-c voltage at the output will change the operation of succeeding equipment.

Radiation then has two effects on electronic equipment: to cause transient amplifier malfunction or to cause catastrophic failure through drastic component change.

The preceding discussion has assumed specific and constant environmental conditions. These conditions are representative only. Each application will present new and different conditions which may vary with time and include other environmental stresses such as temperature extremes.

The problems associated with designing and developing radiation tolerant equipment begin with the environment. The range of dose-rates to be expected and the duration of each must be established or predicted. The equipment life-expectancy should be specified. Maintenance procedures should be described. Only when the environmental specification has been written and the life requirement and maintenance procedure agreed on, can the initial electronic design begin. Material choice is limited by the maximum dose-rate and the life required. System design is restricted by the materials and techniques which can be used.

This design procedure dictates a series of compromises. The cost of a completely inorganic system is exorbitantly high, and it is difficult to obtain good performance while meeting minimum weight and volume requirements. The normal array of components contains materials known to be sensitive. Selection of components constructed of more tolerant materials still does not give unlimited life. Even though a satisfactory life expectancy may be obtained, performance and size may be unsatisfactory.

To demonstrate, again consider the amplifier of Fig. 3. Fig. 7 indicates the probable life expectancy for electronic equipments containing the indicated materials. While is is not possible to predict the interaction of several materials and combined stresses using the available data, the analogy with the weakest link of a chain is applicable. Long life may be realized in many instances, but the weak material will always limit the life of its component.

For operation at a low dose-rate, only a few limitations apply. It may be feasible to use Teflon. As lifetime requirements increase, even at very low dose-rates, the weak materials begin to drop out. If high temperatures are expected life is considerably shortened. At high dose-rate only inorganic systems will give acceptable life-expectancies.

Materials are not the only limitation. Circuit design is restricted. Transient conductivity limits the maximum resistor value which is also restricted by the larger changes of high value resistors.

Fig. 8 is the schematic of an amplifier designed to tolerate high dose-rates for one thousand hours or more.[6] Note the low resistor and capacitor values; 56 kilohms and 0.01 microfarad maximum. There are no organic materials used. The amplifier has a gain of 40 decibels from 10 kc to 1,500 kc over a wide temperature range and will meet military specifications. It is an excellent example of the application of the points presented in this discussion of nuclear radiation effects on electronic components.

Suggestions for designing radiation-tolerant electronic amplifiers are given as follows:

1. Keep resistance values low.
2. Use materials compatible with the environment.
3. Avoid large capacitor values.
4. Keep the circuit simple.
5. Design the system with these restrictions in mind.

References

1. Conductivity Induced in Solid Insulating Materials During Gamma Irradiation, G. C. Huth. *Technical Report no. 58-331*, General Electric Company, Cincinnati, Ohio, 1958.

2. Estimated Radiation Stability of Aircraft Components, C. G. Collins, et al. *APEX 357*, General Electric Company, 1958.

3. The Effect of In-Pile Radiation of the G. E. Ceramic Triode, J. R. Crittenden. *APEX 347*, General Electric Company, 1958.

4. Electron Tubes for Critical Environments, W. H. Kohl, P. Rice. *WADC TR 57-434*, Stanford Research Institute, Stanford, Calif., 1958.

5. Effects of Nuclear Radiation on Electronic Components, E. R. Pfaff, R. D. Shelton. *WADC TR 57-361*, Admiral Corporation, New York, N. Y., 1958.

6. Design and Development of 600° F. Pulse Preamplifiers for Nuclear Environment, W. Friaby, E. Palmer. General Electric Company, Burlington, Vt., May 1959. (Personal Communication)

Reliability Analysis for Aircraft Generators

J. T. DUANE
ASSOCIATE MEMBER AIEE

L. J. YEAGER
ASSOCIATE MEMBER AIEE

EVER-INCREASING reliability wherever economically justifiable long has been an important objective for the designers of all electric equipment. Rough yardsticks—such as service life, maintenance costs, and complaint charges—have been used as measures of success in the attainment of this objective, but precise, quantitative indices of reliability have been lacking in many cases. This lack of emphasis on the quantitative evaluation of product reliability may have been justifiable in the past, but the situation is changing rapidly. In particular, components for aircraft and missile applications will be required more and more frequently to attain specified levels of reliability. This will be necessary because performance requirements in such applications require that components be pushed to the extremes of their capabilities. In addition, the dollar and time investment required to bring a modern, air-borne weapons systems to an operational status, coupled with the vital importance of such systems performing their assigned tasks, makes it imperative that all components not only meet specified performance standards, but that they do so reliably and consistently.

Equipment failures are not individually predictable, but they are subject to analysis using the tools of mathematical statistics. For this reason, statistical techniques play an important part in any study of system or component reliability. However, the problem is not simply a mathematical one. The statistical analysis must be closely coupled with a clear understanding of actual behavior of the equipment to which the reliability analysis is to be applied. The discussion presented here proposes to consider the problem of reliability evaluation in terms of its application to rotating electric machinery used in aircraft and missiles. Particular attention is paid to equipment required to repetitively complete missions of specific length during a relatively long service life. In addition, emphasis is placed on equipments which exhibit wear characteristics such that the likelihood of failure varies with the length of time which the equipment has been in service.

Much has been written in the technical literature concerning reliability, but there still seems to be a need for industry-wide agreement on a precise definition of the meaning of the numerical value of reliability associated with specific pieces of equipment. This problem of definition is considered here in some detail and a specific form of definition is proposed for use.

Definition of Reliability

It is generally agreed that reliability may be defined as the probability that a given component or system will successfully perform some assigned task without departing from predetermined performance limits. Such a definition is general and useful, but it leaves much unsaid. It implies a careful description of the assigned task and the conditions under which it is to be performed. Formulating this sort of description is in itself a major problem. If the assigned task is to be performed repetitively, additional complications arise when the probability of successful completion is not the same on each attempt. It is essential that all such problems be resolved if a number is to represent reliability and mean the same thing to all who use it.

Environment

In any actual application the reliability of a component may be determined by observation of its performance. If a large number of items are observed over a long period of time, sufficient data will be

Paper 59-883, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 25, 1959; made available for printing April 23, 1959.

J. T. Duane and L. J. Yeager are with the General Electric Corporation, Erie, Pa.

The authors would like to express their appreciation to R. G. Blanco and D. A. Wilhelmson for their guidance and assistance in the preparation of this paper and review of the manuscript. They would also like to acknowledge the extensive use of unpublished General Electric Company notes and reports written by R. G. Blanco, P. Gunther, and J. A. Zoellner.

obtained to permit a statistical determination of the component's reliability. The specific manner in which such data may be used to calculate reliability will depend on other facets of the definition, but it will suffice for the moment to state that it can be calculated on the basis of field data. The reliability determined on the basis of actual operating data will be referred to here as "operational reliability." There are other ways in which reliability can be determined and defined, but it is important to remember that operational reliability is the single end result of any worthwhile reliability program. It is the only index that counts once the equipment is in service.

Unfortunately, an accurate determination of operational reliability cannot be made until the equipment has been in service for some time. As weapons systems become more sophisticated, it becomes increasingly important that the organization responsible for the development of the system insure in advance that the completed system will meet a specified value of operational reliability. In order to do this the systems manager requires that the suppliers of subsystems and components deliver products of specified reliability. This is most effectively accomplished by making product reliability a part of the suppliers contractual commitments. The time lag required for the accurate determination of operational reliability makes it an unsatisfactory index for vendor supplied components. Vendors want some means for demonstrating that they have met their commitments and are entitled to payment, while the purchaser wants assurance that the items he buys will do the job expected of them.

The need for an index of reliability prior to operational use leads to the definition of a different type of reliability which will be referred to here as "simulated-service reliability." Such an alternative definition should not be considered as the establishment of a different index of performance, but rather as explicit recognition of the fact that no test program short of actual operational use can completely determine the operational reliability of a product. Simulated-service reliability can be defined as an index of product performance which is the best available approximation to operational reliability that can be determined in a relatively short time using existing (or readily obtainable) test facilities. It is important to remember that this is at best only an approximation to the real thing. It is the responsibility of individuals preparing and issuing specifications on reliability to specify test conditions and techniques that are as close to actual operation conditions as is economically feasible. Only then will simulated-service reliability provide useful predictions of product performance in actual service.

Operating conditions encountered by a product in aircraft service depend on a number of variables. Some of these are independent and some are interrelated, but most can vary over relatively wide ranges. Any single operating condition can be considered as a single point in a mutidimensional stress space where each dimension represents one of the variables or parameters which can affect the performance or life of the equipment in question. Many parameters are required to completely describe any actual operating condition, but it is difficult to visualize any situation involving more than three dimensions. A stress-space diagram for a hypothetical and highly simplified application is shown in Fig. 1. This represents an application where temperature may vary between −50 and 200 C (degrees centigrade), loads between 50 and 100%, and vibration levels between 3 and 6 times the acceleration of gravity. Such an application is therefore represented by the indicated volume on Fig. 1. This alone would not be enough to completely define the application. In addition, a complete description would require a probability density function to define the likelihood of the operating point falling within any incremental element of the indicated volume. This would make a complicated picture by itself, but it is easy to think of additional parameters such as time, vibration frequency, and altitude which should be added to make the representation complete.

Statistical techniques can be developed to analyze the stress-apace representation of an application and relate it to the laboratory performance of a product when subjected to the same stresses. These techniques require the statistical evaluation of all of the interrelationships between the various stresses which may be exerted. Methods of this type would obviously be quite sophisticated and expensive to employ. They may in time be developed to the point where they can be used effectively in evaluating the simulated-service reliability of air-borne rotating machines, but that day does not seem to be here yet.

Historically, specifications for aircraft equipment have been written to require experimental demonstration of ability to perform at a limited number of points in the anticipated stress-space representation of the actual application. The points

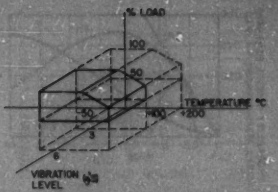


Fig. 1. Simplified stress-space representation of hypothetical aircraft generator environment

selected were those which experience and intuition suggested as the most severe. Some parameters such as vibration, acceleration, fungus, and humidity have been assumed to have effects relatively independent of all the others, and have been tested separately. Experience has shown assumptions of this type to be valid and, in addition, they are frequently necessary because of the limitations of existing test facilities. Specifications of this type have proved practical, and are frequently required by the economics of a given situation. They will probably continue in use.

Simplification and approximation are necessary in order to permit laboratory determination of simulated service reliability. The trend in recent years has been to increase the number of points used to simulate the actual application stress space in the laboratory. This trend probably will and should continue. However, it is important that this not be overdone. Increased complexity of test methods for reliability are justifiable only to the extent that they provide better correlation between simulated service reliability and operational reliability. It is important that adequate records be maintained on both types of reliability so that the correlation between the two may be continually evaluated.

Time Variation of Failure Rate

Exact definition of reliability requires a decision as to whether the numerical value of the probability of successful operation refers to the completion of a single mission (one cycle of operation) or to completion of the total design life. For aircraft and other weapons systems designed for repeated use it seems most logical to specify the probability of completing a single mission. In other words, the most important question is whether or not the system will successfully accomplish its assigned task when called upon to do so. This is the definition chosen for this discussion.

Once reliability is defined in terms of mission completion another question

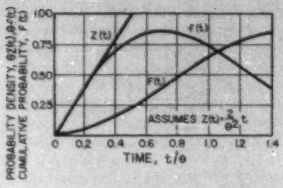


Fig. 2. Typical statistical functions for failure rate analysis

arises. Is the probability of successful mission completion as high for a component nearing the end of its design life as for a brand new unit? The most intuitively appealing answer to this question is negative. Most people feel they take a greater risk of breakdown trouble when they start a trip in a car with 90,000 miles on the speedometer than when they start out in a new car. However, this intuitive approach is not entirely satisfactory. On some types of equipment the probability of failure is no higher for units 1,000 hours old than it is for new units. On other types of equipment the probability of failure increases materially with the age of the item.

Since there is a possibility that the probability of successful mission completion will vary with age, provision should be made for time variation in any complete definition of reliability. This situation makes it desirable to introduce the tools of mathematical statistics. Statistical techniques for the analysis of reliability are available in the literature,[1] and need not be developed here. However, it will be helpful to introduce and define certain functions which are useful in the formulation of exact definitions for reliability.

The first and most familiar function in the analysis of failure phenomena is the probability density function of failure $f(t)$. This function represents the density of the probability of failure as a function of time. The integral of $f(t)$ over any time interval represents the probability of failure within that interval. Since the probability of failure at some time is unity

$$\int_0^\infty f(t)dt = 1.0 \tag{1}$$

The cumulative distribution function of failures $F(t)$, represents the probability of failure at some time prior to the time t. The cumulative distribution may be obtained by integration of $f(t)$

$$F(t) = \int_0^t f(t)dt \tag{2}$$

The probability of survival to time t is denoted $G(t)$ and may be expressed as

$$G(t) = 1 - F(t) \tag{3}$$

One other function which will prove useful later is the conditional probability density of failure $z(t)$. This is the probability density of failure at time t, given the condition that the item is alive at t

$$z(t) = \frac{f(t)}{G(t)} \tag{4}$$

Complete specification of any one of these functions, $f(t)$, $F(t)$, or $z(t)$, permits the determination of all the others. Fig. 2 pictures $f(t)$, $F(t)$, and $z(t)$ for the special case where $z(t)$ is a linear function of time. The expected life of the item described by these curves is equal to the mean of $f(t)$ or the expected time to failure T_m

$$T_m = E[t] = \int_0^\infty tf(t)dt \tag{5}$$

It has already been stated that reliability is to be defined as the probability of successfully completing a single mission. This necessarily assumes that the item in question be alive at the beginning of the mission. If the mission is h hours long and begins at time t, the probability of the item having failed at some time prior to the end of the mission will be $F(t+h)$. The probability of its having failed prior to the start of the mission is $F(t)$ so that the probability of an item failing in the h hours starting at time t will be $F(t+h)-F(t)$. This probability, divided by the probability of being alive at the start of the mission, yields the probability that the item will fail during the mission on the condition that it is alive at the start. For example, if only one-third of the original units are alive at t, then an expected mortality of 10% of the original units is equivalent to 30% of the survivors

$$F(h|t) = \frac{F(t+h)-F(t)}{G(t)} = \frac{F(t+h)-F(t)}{1-F(t)} \tag{6}$$

The notation $F(h|t)$ is to be taken as the probability of h (failure during the next h hours) on the condition t (being alive at time t).

It is now possible to define the reliability $r(h|t)$ as the probability of completing an h hour mission if alive at t

$$r(h|t) = 1 - F(h|t) = \frac{1-F(t+h)}{1-F(t)} \tag{7A}$$

$$r(h|t) = \frac{G(t+h)}{G(t)} \tag{7B}$$

At this point it would be possible to select as a definition of reliability the minimum value which r will be allowed to assume during the design life of the equipment. For many, but by no means all, equipments r will be a minimum when t is maximum. In such cases reliability will be at a minimum when the equipment has completed its design life and is ready for removal for overhaul or replacement. If the scheduled operating time to replacement is T_r, reliability might be defined as

$$R_d(h|T_r) = \frac{G(T_r+h)}{G(T_r)} \tag{8}$$

While this R_d reliability is a possible choice for an exact definition, it does have some drawbacks. These can best be seen by considering the use to which component reliability data will be put. The primary purpose for establishing component reliability is to insure an acceptable system reliability. If no redundancy exists in the system, its reliability will be equal to the product of the reliability values for the components. If we denote the probability of successful completion of an h hour mission for the system, as $r(h|t)$, and the reliability of the ith component as $r_i\ (h|t-T_i)$, system reliability may be written as

$$r(h|t) = r_1(h|t-T_1)r_2(h|t-T_2)\dots r_s(h|t-T_s) \tag{9A}$$

$$r(h|t) = \prod_{i=1}^{s} r_i(h|t-T_i) \tag{9B}$$

In equation 9, s is the number of components in the system and T_i is the time of the most recent installation of a new item for the ith component. Inclusion of T_i makes the expression general enough to allow for continued maintenance and replacement.

It should be apparent from equation 9(B) that the system reliability may vary with time. For the special case where all s components have minimum reliability at maximum age ($t-T_i=T_{ri}$), and maximum reliability at $t=T_i$ the minimum and maximum values of r may be established as

$$r_{\min} = \prod_{i=1}^{s} r_i(h|T_{ri}) \tag{10A}$$

$$r_{\max} = \prod_{i=1}^{s} r_i(h|0) \tag{10B}$$

For any system with a fixed replacement or maintenance schedule for all components there will exist minimum and maximum values of r. There is some possibility that r will take on values anywhere between its minimum and maximum values. This situation is pictured in Fig. 3 which shows a plot of a possible probability density function of r denoted as $p(r)$. It seems reasonable to expect that for systems with large numbers of components, $p(r)$ is not apt to be high

for values of r very far removed from some mean value. Even without proof of this assumption, it is possible to state that the most likely or expected value of r will be represented by the first moment of the function $p(r)$. If reliability is to be defined for a system used in large numbers with each system including many components, the expected value of r will give an excellent approximation to the reliability of a single system selected at random from the group at any arbitrary time.

One of the theorems of mathematical statistics states that the expected value of a function which is equal to the product of several other independent functions is the product of the expected values of the other functions

$$E[r] = R_s \tag{11}$$

$$R_s = \prod_{i=1}^{n} E[r_i(h|t-T_i)] \tag{12}$$

Thus, it seems reasonable to define the reliability for any single component as the expected value of its instantaneous reliability. Expected value for component reliability may be evaluated in the following manner.

Consider a situation where a single component in a given system is observed through N replacements. The data obtained in such an observation will be statistically equivalent to a situation where N components are all placed on test at the same time. Since some of the components will fail during the test, more component-hours will be accumulated during earlier portions of the service life. This will weight the expected reliability in favor of "low-time" reliabilities. The relative weight (probability density) of each value of reliability is just the number of components still operating at the time associated with that reliability, divided by the total component-hours accumulated during the test. For this type of test, the number of components alive at any time t will be $NG(t)$ as long as $t<T_r$. The reliability for any component will be given by equation 7(B) as a function of time. At any time $t(0<t<T_r)$ the instantaneous component reliability will be $r(h|t)$ and the probability density of observing this reliability will be

$$p(r) = \frac{NG(t)dt}{\int_0^{T_r} NG(t)dt} \frac{1}{dt} \tag{13}$$

The product of $p(r)$ and r integrated over the limits $-\infty$ to $+\infty$ is equal to the first moment of $p(r)$ or the expected value of r. For the situation considered here $p(r)$ is zero for $t<0$ and $t>T_r$ so that it is possible to make a change in variable and integrate over the limits $t=0$ to $t=T_r$. Thus the expected value of instantaneous reliability R_s for a component will be

$$R_s = \int_0^{T_r} r_t(h|t) \frac{G(t)}{\int_0^{T_r} G(t)dt} dt \tag{14}$$

but

$$r(h|t) = \frac{G(t+h)}{G(t)} \tag{7B}$$

and

$$R_s = \frac{\int_0^{T_r} G(t+h)dt}{\int_0^{T_r} G(t)dt} \tag{15}$$

The development of equation 15 was in terms of a system with no redundant components. If system failure requires the failure of two or more components, the system reliability as given by equations 9 will involve some sums of component reliability values. The same analysis will still apply since the expected value of a function equal to the sum of several other independent functions will be the sum of the expected values of the other functions.

Equation 15 is a mathematical statement of an exact definition of reliability which states that reliability is the mean value observed in service (or on test) of successfully completing a single mission. In the authors' opinion, this R_s reliabiliiy is the most realistic definition to use for the specification of component reliability. It is preferable to the minimum or R_d reliability. It should be noted that if the failure characteristics (the form of the $z(t)$ function) are known, equivalent performance may be obtained by specifying either R_d or R_s, but the numerical values of the two will frequently be different.

When operational reliability is considered, the superiority of the R_s definition becomes more apparent. To determine the R_s value of operational reliability it is necessary to observe a number of systems in operation over a period of many times the replacement life of the component in question. If each mission observed is classed as either a successful one or a failure, the R_s value of operational reliability will be simply the ratio of successful missions to total missions. A similar determination of R_d value of operational reliability would necessitate first determining at what component age the highest percentage of components (of that age) were failing and then calculating the R_d reliability as the ratio of successful missions to total missious for components of that age.

Bither the operational or simulated service values of reliability as defined here will depend not only on the failure characteristics of the equipment and the mission life but also the scheduled replacement time. Thus maintenance and replacement policy may have an important effect on the reliability obtained from a given component. For some items it may be possible to set up inspection procedures which will warn of incipient deterioration and in effect insure as-good-as-new condition for a component which has been in service for some time. Periodic inspection in this case may have somewhat the same effect as reducing replacement life. All of these factors can be important in establishing operational reliability and it is therefore important that they be fully considered in specifying programs for the measurement of simulated-service reliability.

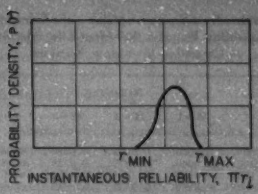


Fig. 3. Typical probability density distribution of system reliability

Wear-Out Phenomena

Before considering methods for the measurement of simulated-service reliability, it will be helpful to consider in some detail the nature of the time dependence of failure rates. Much work has been done on the analysis of reliability problems for situations where failures are randomly distributed. The assumption of random failure distribution has been justified on the basis of statements that any other types of failure are due to deficiencies in the design or application of the system or its components.[1] A truly random failure distribution presupposes the absence of any significant fatigue or wearout effects. While this may be a valid assumption in some situations, the authors do not feel that it is appropriate for a general analysis of rotating electrical machinery reliability.

Analysis of various failure rates is most easily accomplished by use of the conditional probability density of failure $z(t)$. The other distribution functions of equations 1, 2, and 3 may be rewritten in terms of $z(t)$ by recognizing that

$$f(t)dt = -d[G(t)] \tag{16}$$

and from equation 4

Table I. Minimum Value of Generator Reliability R_{dw} of 1,000-Hour Life Machine Operating in 4-Hour Missions

θ	Reliability, R_{dw}			
	M=1	M=2	M=3	M=4
1,000	0.9880	0.9931	0.9880	0.9840
1,100	0.9903	0.9935	0.9910	0.9890
1,200	0.9908	0.9945	0.9929	0.9922
1,300	0.9909	0.9932	0.9944	0.9944
1,400	0.9972	0.9959	0.9955	0.9953
1,500	0.9974	0.9964	0.9964	0.9968
1,600	0.9975	0.9968	0.9970	0.9976
1,700	0.9977	0.9972	0.9975	0.9981
1,800	0.9978	0.9975	0.9979	0.9983
1,900	0.9979	0.9978	0.9983	0.9988
2,000	0.9980	0.9980	0.9985	0.9990

$$\int_0^t z(t)dt = -\ln G(t) \tag{17}$$

$$G(t) = \exp\left[-\int_0^t z(t)dt\right] \tag{18A}$$

$$f(t) = z(t)\exp\left[-\int_0^t z(t)dt\right] \tag{18B}$$

$$F(t) = 1-\exp\left[-\int_0^t z(t)dt\right] \tag{18C}$$

In addition, the function $z(t)$ is useful because it may be conveniently determined from life-test data. Differentiation of equation 17 gives

$$z(t) = -\frac{d[\ln G(t)]}{dt} \tag{19}$$

or

$$z\left(t+\frac{\Delta t}{2}\right) \cong \frac{\ln G(t) - \ln G(t+\Delta t)}{\Delta t} \tag{20}$$

If a large sample of components is placed on test at time $t=0$, and $n_s(t)$ represents the number of units still alive at time t

$$G(t) = \frac{n_s(t)}{n_s(0)} \tag{21}$$

and

$$z\left(t+\frac{\Delta t}{2}\right) \cong \frac{\ln n_s(t) - \ln n_s(t+\Delta t)}{\Delta t} \tag{22}$$

Given a sufficiently large sample, the function $z(t)$ may be plotted from life-test data using equation 22.

In the analysis of life-test data it is frequently helpful to fit the observed data to some statistical distribution function which can be conveniently treated by analytical methods. Many different functions have been used for this purpose, but one of the most general and widely applicable was proposed by Weibull.[8] The Weibull distribution is particularly useful for failures caused by fatigue or wearout effects. It may be written as

$$z_w(t) = \frac{m}{\theta^m} t^{m-1} \tag{23}$$

Substitution into equations 18 yields

$$G_w(t) = \exp\left[-\left(\frac{t}{\theta}\right)^m\right] \tag{24A}$$

$$f_w(t) = \frac{m}{\theta^m} t^{m-1}\exp\left[-\left(\frac{t}{\theta}\right)^m\right] \tag{24B}$$

$$F_w(t) = 1-\exp\left[-\left(\frac{t}{\theta}\right)^m\right] \tag{24C}$$

The mean life of components subject to Weibull distributed failures (T_{mw}), may be calculated as

$$T_{mw} = \int_0^\infty t f_w(t)dt \tag{25A}$$

$$T_{mw} = \theta\left(\frac{1}{m}\right)! \tag{25B}$$

If failures are distributed in accordance with Weibull's function, the minimum R_d, and expected, R_e, values of reliability may be written with the aid of equations 8 and 15 for the case $m \geq 1$:

$$R_{dw}(h \mid T_r) = \exp\left[\frac{T_r^m - (T_r+h)^m}{\theta^m}\right] \tag{26}$$

$$R_{ew} = \frac{\int_0^{T_r} \exp\left[\left(\frac{t+h}{\theta}\right)^m\right]dt}{\int_0^{T_r} \exp\left[\left(\frac{t}{\theta}\right)^m\right]dt} \tag{27}$$

where the subscript w implies restriction to the Weibull distribution. While equation 26 may be solved directly for θ when R_{dw}, T_r, h, and m are known, equation 27 presents more of a problem and in general can be best handled by trying values of θ and solving for R_{ew}. In either case, numerical solutions may be readily obtained with today's digital computing equipment. As an example, Table I shows values of R_{dw} obtained for various values of θ and m when the application requires 1,000 hour life equipment to perform successive 4-hour missions. Table II shows similar results for the R_{ew} reliability.

In the special case where $m=1$, the Weibull distribution reduces to the well-known exponential distribution which describes the situation where failures are randomly distributed. In this case the mean life is also the mean-time-between-failures and is equal to θ. Also, since failure rate is not time dependent, the R_d and R_e reliability values are identical.

The Weibull distribution is important in the analysis of reliability for rotating machinery because it provides a good description for the failure of machine components. Failure rates for bearings, brushes, insulation systems and structural parts can all usually be correlated with the Weibull distribution. Fig. 4 shows results obtained by testing to failure (under standard conditions) 40 samples of a single insulation system. These data were analyzed in accordance with equation 22 and the resulting $z(t)$ plotted in a non-dimensionalized form. In this case the data suggest a Weibull failure distribution with $m=3$.

Table II. Expected Value of Generator Reliability R_{ew} of 1,000-Hour Life Machine Operating in 4-Hour Missions

θ	Reliability, R_{ew}			
	M=1	M=2	M=3	M=4
1,000	0.9960	0.9966	0.9969	0.9970
1,100	0.9964	0.9971	0.9975	0.9978
1,200	0.9967	0.9975	0.9980	0.9983
1,300	0.9969	0.9979	0.9984	0.9987
1,400	0.9971	0.9981	0.9987	0.9990
1,500	0.9973	0.9983	0.9989	0.9992
1,600	0.9975	0.9985	0.9991	0.9994
1,700	0.9976	0.9987	0.9992	0.9995
1,800	0.9978	0.9988	0.9993	0.9996
1,900	0.9979	0.9989	0.9994	0.9997
2,000	0.9980	0.9990	0.9995	0.9998

In spite of the fact that most generator components tend to exhibit Weibull distributed failures with $m>1$, there is no assurance that failure rates for complete generators will be Weibull distributed. The act of combining components into a generator system introduces the possibility of additional failures due to interactions which may be random in nature. These would correspond to the random failures discussed by Wilson.[2]

One excellent source of failure rate information is field service experience. Fig. 5 shows plots of $z(t)$ for four different types of aircraft generators and starters manufactured by the authors' company. Failure data was obtained from field reports and analyzed by the methods of equation 22. Each curve represents a class of machines and includes more than one model. The a-c generator curve of Fig. 5 did exhibit a slight "infant mortality" effect, but the infant mortality period was eliminated from the analysis to provide a consistent picture of the initial failure rate effects without consideration of infant mortality. Failure data used

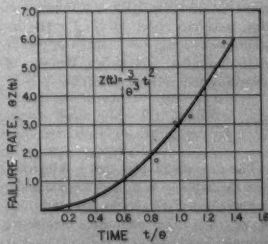


Fig. 4. Failure rate analysis based on laboratory insulation system tests

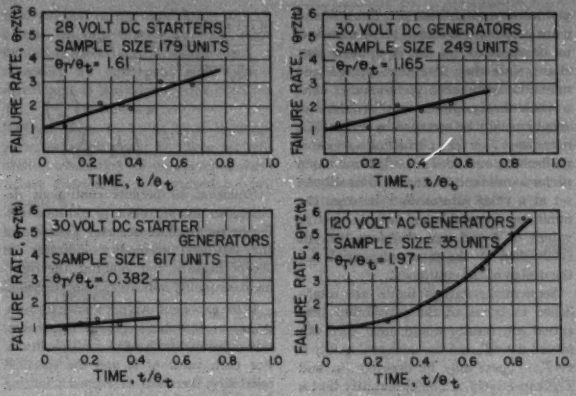


Fig. 5. Field service data on generator failure rates

to prepare Fig. 5 included machine removals during ground check-out and inspection and therefore the curves are not suitable for numerical evaluation of in-flight reliability. However, they do accurately portray the time-dependent nature of $z(t)$.

Particularly significant in Fig. 5 is the fact that in no case did the equipment show a zero value for $z(0)$, and in each case $z(t)$ was in increasing function of time. The Weibull distribution is nonzero at $t=0$ only for the case $m=1$, and in this case $z(t)$ does not increase with time. Since the data of Fig. 5 can not be satisfactorily correlated with the Weibull distribution, some other distribution is needed. A convenient solution is offered by Flehinger and Lewis.[1] They suggest a distribution of the form

$$z_f(t)=\frac{1}{\theta_r}+\frac{m}{\theta_t^m}t^{m-1} \tag{28}$$

This can be recognized as a combination of two different modes of the Weibull function. The other pertinent distributions can be written

$$G_f(t)=\exp\left[-\frac{t}{\theta_r}-\left(\frac{t}{\theta_t}\right)^m\right] \tag{29A}$$

$$f_f(t)=\left[\frac{1}{\theta_r}+\frac{m}{\theta_t^m}t^{m-1}\right]\exp\left[-\frac{t}{\theta_r}-\left(\frac{t}{\theta_t}\right)^m\right] \tag{29B}$$

$$F_f(t)=1-\exp\left[-\frac{t}{\theta_r}-\left(\frac{t}{\theta_t}\right)^m\right] \tag{29C}$$

Mean life for items subject to this z_f failure distribution are given as[1]

$$T_{mf}=\int_0^\infty \exp\left[-\frac{t}{\theta_r}-\left(\frac{t}{\theta_t}\right)^m\right]dt \tag{30}$$

Closed form evaluation of equation 30 is not practical, but it can be evaluated numerically when the parameters θ_r, θ_t, and m are specified. Fig. 6 shows the ratio T_{mf}/θ_t as a function of θ_t/θ_r for the case $m=2$. Note that this distribution also reduces to the exponential distribution for the special case $m=1$.

The z_f failure distribution leads to reliability expressions similar to equations 26 and 27 if $m>1$

$$R_{sf}(h\mid T_r)=\exp\left[\frac{T_r^m-(T_r+h)^m}{\theta_t^m}-\frac{h}{\theta_r}\right] \tag{31}$$

$$R_{sf}=\frac{\displaystyle\int_0^{T_r}\exp\left[-\frac{t+h}{\theta_r}-\left(\frac{t+h}{\theta_t}\right)^m\right]dt}{\displaystyle\int_0^{T_r}\exp\left[-\frac{t}{\theta_r}-\left(\frac{t}{\theta_t}\right)^m\right]dt} \tag{32}$$

Again, equations 31 and 32 are best evaluated numerically. Fig. 7 shows a graphical representation of the relationship between R_{sf} and the parameters θ_r and θ_t for equipment with a 1,000-hour replacement life and a 4-hour mission life.

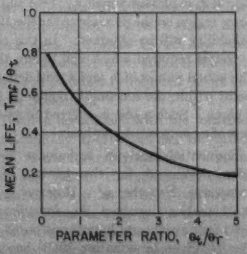


Fig. 6. Mean life determination for Flehinger-Lewis distribution failures

Reliability Testing

Determination of simulated-service reliability requires some measuring technique which is presumably based on the testing of samples selected from a large population. To the extent that operational environment can be simulated, a large sample could be set in operation and observed over a long enough period of time to satisfy all interested parties that the ratio of successful operational cycles to total attempted cycles provides a valid measure of reliability. This would probably be an inelegant, uneconomical, and inaccurate approach to the problem. Fortunately, statistical theory provides more efficient test methods.

Maximum efficiency in testing of this type will be obtained by the use of the techniques of sequential analysis. The theory of sequential analysis is available elsewhere,[4] and need not be repeated here. It will suffice to point out that sequential analysis provides a method for determining whether or not one or more parameters of a distribution function describing a population lie within acceptable limits. For reliability testing, the parameters to be evaluated will be the coefficients and exponents in the functional representation of $z(t)$. It is necessary that the form of $z(t)$ be known or assumed, but it is possible to assume forms of wide generality. The general procedure for determination of reliability will be as follows if the form of $z(t)$ is known.

1. From the specified reliability figure determine the extreme values which the unknown parameters in the function $z(t)$ may assume without resulting in a reliability below the specified value.

2. Design and conduct a test to insure within a specified confidence level that each of the unknown parameters of $z(t)$ lies in a region acceptable by the standards established in step 1.

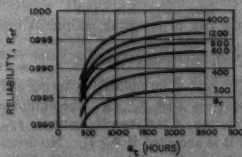


Fig. 7. Failure parameter determination for R_{sf} reliability where mission length is 4 hours, service life is 1,000 hours, and $m=2$

For most generator components being tested separately under controlled conditions, the Weibull distribution may be assumed. In these cases, the nature of failure mechanisms will usually permit predetermination of the value of the exponent m. In such cases, the test procedure need consider only a single parameter.

Single Parameter Tests for Weibull Distribution

General techniques for the sequential analysis of test results are described by Wald,[4] and the specific example of a sequential analysis procedure for the Weibull distribution with $m=1$ is given by Wilson.[5] This portion of the discussion considers the more general case of sequential analysis testing equipment known to have Weibull distributed failures where the exponent m is known but may have any real value. It is important to note that the test procedure can only determine that θ (in the function of equation 23) is not less than some limiting value T_1. The numerical value of this limit on θ must be established for a given reliability by a separate analysis using equation 27 or Table II.

Operating environment and the actual mechanics of equipment operation during the test must also be established separately. Since the test considered here is intended to evaluate the equipment's failure rate throughout the entire service life, expected mission length will not necessarily be an important factor in establishing test conditions. In addition, an exact definition of what constitutes failure must be established. Once these steps have been taken, sequential analysis provides ground rules for determining whether the test should continue or be stopped with either acceptance or rejection of the equipment on test. In order to formulate the criteria for acceptance or rejection, four numbers must be established:

α = the risk of rejecting equipment which is actually acceptable, i.e., has higher θ than allowable minimum
T_0 = some value of θ higher than the minimum acceptable value
β = the risk of accepting equipment which should actually be rejected
T_1 = the minimum acceptable value of θ

The test is formulated in terms of the hypothesis that $\theta \geq T_0$. The equipment is rejected if test results indicate $\theta \leq T_1$ and it is accepted if $\theta \geq T_0$. No decision can be made as long as results indicate $T_1 < \theta < T_0$. Minimum acceptable reliability determines the value of T_1 and the allowable risk of accepting substandard performance establishes β. Thus, T_1 and β are properly considered test parameters to be specified by the equipment purchaser since they determine his risk and minimum equipment performance. The α risk is the equipment suppliers risk and should usually be specified by him. The parameter T_0 is set up to define a region of no decision and does not affect minimum performance limits. It should usually be set at a value intermediate between the minimum acceptable and design objective values of θ and can therefore best be established by the group responsible for the design of the equipment.

Sequential analysis limits are established in terms of the probability that actual test results would have been observed if $\theta = T_0$ and the same probability if $\theta = T_1$. These may be denoted P_{0n} and P_{1n} respectively. The probability that n failures will occur at the times $t_1, t_2, \ldots t_n$ if $\theta = T_0$, with $f(t)$ given by equation 24(B) is P_{0n}

$$P_{0n} = \prod_{i=1}^{n} f(t_i, T_0) \tag{33A}$$

$$P_{0n} = \prod_{i=1}^{n} \frac{m}{T_0^m} t_i^{m-1} \exp\left[-\left(\frac{t_i}{T_0}\right)^m\right] \tag{33B}$$

$$P_{1n} = \prod_{i=1}^{n} \frac{m}{T_1^m} t_i^{m-1} \exp\left[-\left(\frac{t_i}{T_1}\right)^m\right] \tag{33C}$$

$$\frac{P_{1n}}{P_{0n}} = \left(\frac{T_0}{T_1}\right)^{nm} \prod_{i=1}^{n} \exp\left[t_i^m\left(\frac{1}{T_0^m} - \frac{1}{T_1^m}\right)\right] \tag{34}$$

Derivation of test limits is given by Wald.[4] If

$$B < \frac{P_{1n}}{P_{0n}} < A \tag{35}$$

the test is continued with no decision. If

$$\frac{P_{1n}}{P_{0n}} \geq A \tag{36}$$

the test is terminated with rejection of the equipment and the conclusion $\theta < T_1$. If

$$\frac{P_{1n}}{P_{0n}} \leq B \tag{37}$$

the test is terminated with acceptance and the conclusion that $\theta > T_0$. The limits A and B are fixed by the risks α and β

$$A = \frac{1-\beta}{\alpha} \tag{38A}$$

$$B = \frac{\beta}{1-\alpha} \tag{38B}$$

Substituting equations 33 and 34 into 35, taking the natural logarithm and rearranging puts the inequality in a more convenient form

$$\frac{\ln B + nm \ln\left(\frac{T_1}{T_0}\right)}{\left(\frac{1}{T_0^m} - \frac{1}{T_1^m}\right)} > \sum_{i=1}^{n} t_i^m > \frac{\ln A + nm \ln\left(\frac{T_1}{T_0}\right)}{\left(\frac{1}{T_0^m} - \frac{1}{T_1^m}\right)} \tag{39}$$

Equation 39 permits continuous decisions to accept, reject, or continue testing as each failure occurs. Violation of the left-hand inequality indicates an accept decision and violating the right-hand inequality indicates a reject decision.

If the exponent m is not known for sure, it may be desirable to provide for evaluation of results in terms of multiple limits. For each value of m and the specified reliability, there will correspond a limiting value of θ. Thus the same set of test results may be evaluated for several values of m with a requirement that acceptance be indicated in each case before the equipment is finally accepted.

Single-parameter sequential-analysis testing can be extremely useful in the evaluation of motor and generator components. However, the data presented in Fig. 5 show that complete machines do not conform to the Weibull distribution. In order to accurately measure reliability for complete machines, some more sophisticated techniques will be required.

Two Parameter Testing

It has already been pointed out that actual machines conform to the more complex distribution used by Flehinger and Lewis (equation 28). In this case also, it seems reasonable to assume that the value of m can be predetermined. If this is done, there are still two parameters which must be considered (θ_r and θ_t). Equation 32 or Fig. 7 may be used to translate specified reliability into a set of allowable values for the coefficients θ_r and θ_t. If θ_r is plotted versus θ_t, there will be some region on the plot where the specified reliability will be obtained. In this case a test is required which will assure that the actual values of θ_r and θ_t will define a point within the acceptable region on the θ plane. Such a test is referred to as the test of a composite hypothesis.

Sequential analysis techniques have been developed for the testing of a composite hypothesis.[4] However, the mathematics involved in formulating tests for a composite hypothesis based on the $f(t)$ given by equation 29(B) become very complicated. At the present time, the authors are unable to offer any prac-

tical solution to the formulation of such a test. In spite of this situation, some method is needed for the measurement of reliability for equipment with failures distributed in accordance with equation 28.

There seem to be at least four approaches which might be taken for reliability evaluation of equipment where failure distribution is governed by two parameters.

1. Develop the mathematical techniques and tabulations of appropriate functions to permit the use of conventional sequential analysis methods for the tests of the composite hypothesis.

2. Arbitrarily assume that the failure distribution conforms to some simpler function such as the Weibull distribution with $m=1$ or 2.

3. Assume that the coefficient θ_f is based on random effects which are not influenced by design and may be considered fixed and known for any given class of equipment. This assumption reduces the problem to a single parameter test to determine the value of θ_t.

4. Assume that the coefficient θ_t can be accurately determined by separate tests on machine components (bearings, insulation, etc.) under ideal conditions. The test problem would then reduce to a single parameter test to measure the random failure rate contribution due to θ_f.

The first of the foregoing approaches is probably the best, but it can not be counted on until it is proved practical. The second is an undesirable but immediately available solution which has already seen some use and will probably see much more. The third approach is equally unattractive and it would be very difficult to justify on a logical basis. Number four is probably the most desirable alternative to the first approach. It would seem to provide for the full consideration of fatigue or aging effects and yet still permit measurement of any random failure phenomena which may be present. Selection of an efficient and practical method for measuring two parameter failure rates to determine reliability (approach number 1) is not a simple task. There does not seem to be any good method available now, and considerable effort will probably be required before one can be developed.

Tests in accordance with the fourth of the foregoing approaches can be formulated in the following manner. Assume generator failure characteristics can be described by $z_f(t)$ as given by equation 28 and that separate tests on the component parts of the generator have established the values of m and θ_t. Let the value of θ_t so determined be T_t. From equation 32 or Fig. 7 determine the minimum value of θ_f which will give the necessary reliability and call this T_1. At the same time select values for α, β, and T_0 as described for single parameter tests (where T_0 is greater than T_1 but less than the design objective value of θ_f). Since this is actually a single parameter test, the limits of equations 35, 36, and 37 will still apply, but the ratio P_{1n}/P_{0n} will be more difficult to handle. However, the ratio can be written using equation 29(B)

$$\frac{P_{1n}}{P_{0n}} = \frac{P_{1(n-1)}}{P_{0(n-1)}} \frac{f_f(T_1, T_t, t_n)}{f_f(T_0, T_t, t_n)} \tag{40A}$$

$$\frac{P_{1n}}{P_{0n}} = \frac{P_{1(n-1)}}{P_{0(n-1)}} \left[\frac{\frac{1}{T_1} + \frac{m}{T_t^m} t_n^{m-1}}{\frac{1}{T_0} + \frac{m}{T_t^m} t_n^{m-1}} \right] \times \exp\left[t_n \left(\frac{1}{T_0} - \frac{1}{T_1} \right) \right] \tag{40B}$$

Thus, if a number of units are placed on test; accept, reject, or continue decisions can be made each time a failure occurs. Each new failure becomes the nth failure and the accumulated operating time on the failed unit is t_n. Using t_n and the previous test results, equation 40(B) can be used to determine P_{1n}/P_{0n}. Equations 35, 36, and 37 provide the basis for a decision.

Predetermination of test time required to evaluate equipment in accordance with equation 40(B) is not a simple problem. Considerable work will have to be done before test organizations can do an efficient job of scheduling and cost estimating for such tests. Some comfort may be gained from the fact that if the anticipated value of θ_f is much lower than T_1, a convenient approximation results from considering T_1 to be infinite. In such a case, tests of the ratio given by equation 40(B) will proceed much as if they were determined by the inequality of equation 39 (with $m=1$). Thus, some insight may be gained as to the most effective scheduling for, and probable duration of the test program.

Up to this point, infant mortality (a high, but rapidly decaying initial value of $z(t)$) has not been included in this analysis. This has been partially because of a desire to limit the length of the discussion, but more importantly it is due to the authors' opinion that it is conservative to ignore infant mortality in the definition and measurement of reliability. It is felt that extra, early failures will result in a greater reduction in the R_s value of simulated-service reliability than in the R_o value of operational reliability. Infant failures will automatically be included in any measurement of operational reliability and they must be fully considered in the equipment manufacturers' design and inspection testing. However, since they occur early in the service life, they will have less effect on operational reliability than on the value determined by simulated life test.

Conclusions

Measurement and analysis of equipment reliability is a broad and complex field. Although much work has already been done, it still represents an extremely fertile field for future work. Of particular importance in providing a sound basis for future reliability work is the accumulation of extensive failure data on systems, machines and machine components. A fund of records permitting the determination of both operational and simulated-service reliability for many different items should prove invaluable in establishing better test methods and conditions.

This discussion has dealt primarily with problems in the definition and measurement of reliability. In these areas several specific conclusions are drawn:

1. Operational reliability and simulated-service reliability are two different things and should be clearly recognized as such. They should be as close together as possible, and the responsibility for the degree of agreement between them must lie with the specification which establishes the methods of measurement.

2. Reliability for an item which must operate for several missions should be defined as the expected value of the probability that it will successfully complete a randomly selected mission (R_s).

3. Typical aircraft rotating machines exhibit a conditional probability density of failure which increases with time and is nonzero at $t=0$.

4. Efficient and economical test methods are available for reliability testing of items (such as machine components) which are known to follow single parameter failure laws. Considerable work will be required to develop equally effective methods for testing complete machines where two parameters are of importance.

5. Until efficient test procedures are available for evaluation of the composite hypothesis required by two parameter failure distributions (z_f), an alternative test method is needed. In this situation the recommended approach is to measure fatigue and wear-out effects by separate component tests and use this information to establish single parameter sequential tests on the complete machine to measure the random failure rate θ_f.

Nomenclature

A = upper limit for probability ratio in sequential analysis

B = lower limit for probability ratio in sequential analysis
E = expected value of a function
F = probability of failure
G = probability of survival
R = reliability or probability of successful performance
T = time (fixed value)
f = probability density of failure
m = exponent in failure distribution
n = number of samples or number of failures in a test
p = probability density (general)
r = reliability (instantaneous value)
s = number of components in a system
t = time (variable)
z = conditional probability density of failure
α = risk of rejecting good equipment
β = risk of accepting inadequate equipment
θ = parameter in failure distribution

SUBSCRIPTS

0 = design value (greater than acceptable minimum)
1 = specified minimum
d = minimum value
e = expected value
f = arising from Flehinger-Lewis failure distribution
m = mean value
r = coefficient associated with random failure rate or removal time
s = survivors
t = coefficient associated with time varying rate
w = associated with Weibull failure distribution

References

1. TWO-PARAMETER LIFETIME DISTRIBUTIONS FOR RELIABILITY STUDIES OF RENEWAL PROCESSES. G. J. Flehinger, P. A. Lewis. *Journal of Research and Development*, International Business Machines Corporation, New York, N. Y., vol. 3, no. 1, Jan. 1959, pp. 58–73.

2. RELIABILITY EVALUATION OF AIRCRAFT AND MISSILE EQUIPMENT, I. THEORY OF SMALL SAMPLE METHODS, B. J. Wilson. *AIEE CP 58-913*, 1958.

3. A STATISTICAL DISTRIBUTION FUNCTION OF WIDE APPLICABILITY, W. Weibull. *Journal of Applied Mechanics*, American Society of Mechanical Engineers, New York, N. Y., vol. 18, 1951, pp. 293–97.

4. SEQUENTIAL ANALYSIS (book), A. Wald. John Wiley & Sons, Inc., New York, N. Y., 1947.

Computer Evaluation of High-Temperature Aircraft A-C Electrical System Designs

W. E. SOLLECITO
ASSOCIATE MEMBER AIEE

D. A. SWANN
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a design and analog computer study of a high-temperature aircraft a-c electrical system. The study includes an extensive, nonlinear representation of the a-c alternator and shows the effects of temperature, load, and regulator and exciter nonlinear behavior on system stability and transient performance. Torque-limiting techniques are evaluated and a computer approach for study of the reactive load division loop is developed.

REGULATION of alternator voltage by means of feedback techniques has been practiced successfully for many years.[1–7] Basically, such systems consist of an alternator, an exciter, a reference source, and regulating networks as shown in Fig. 1. The alternator supplies the a-c electric power, the exciter provides the control signal power level, the reference sets the output voltage level, and the regulating networks provide system stability and dynamic control.

Compared with currently available equipment, early components were bulky, heavy, low in power level, and slow in speed of response. As the components improved, so did system performance. Availability of good low-temperature (−65 to +250 F [degree Fahrenheit]) components and subcomponents allows system designs which are attractive from a size, weight, cost, and performance viewpoint.

For high-temperature systems (−65 to +775 F), the picture changes greatly. Components and subcomponents are either in the development or predevelopment stage, the expected size, weight, and cost are high, the expected range of subcomponent values is small, and the performance characteristics are limited. From a systems point of view the net effect is threefold:

1. Because of size, weight, and power requirements, component higher frequency capabilities are reduced and single-ended, saturated gain characteristics are imposed. These tend to degrade system response time.

2. Available compensation network time constants are presently lower-limited in the 0.5-second region which tends to limit the system loop gain and thereby tends to degrade system accuracy.

3. The large change in temperature causes a correspondingly large change in component parameters which increases the system stability problem.

The design approach must necessarily be one which tends to minimize the deleterious effects of parameter variations and provides a stable system whose dynamic performance is commensurate with the availability of equipment and the size and weight requirements of the system.

System Design

In designing this system, an approximate linear analysis was used to indicate favorable design areas which were then analyzed in a more exact, nonlinear fashion by means of the analog computer. The extensive alternator equations used are shown in Appendix I, the computer setup for this study is described in Appendix II, and the alternator linearization equations are shown in Appendix III.

The over-all system performance speci-

Paper 59-907, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing May 15, 1959.

W. E. SOLLECITO and D. A. SWANN are with the General Electric Company, Schenectady, N. Y.

The authors wish to acknowledge the assistance and guidance given to this project by their associate engineers. In particular, they express appreciation to Dr. R. T. Smith for an understanding of machine behavior, F. P. deMello for an understanding of the reactive load division problems in parallel operation, Dr. K. G. Black for numerous consultations, R. E. Anderson for equipment analysis, and F. J. Ellert for detailed calculations.

Fig. 1. Basic electrical system

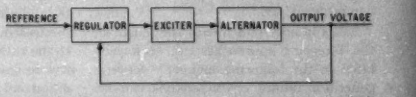
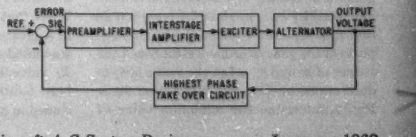


Fig. 2. Amplification stages required to supply appropriate power levels



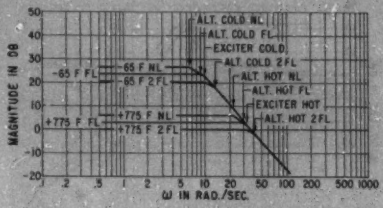


Fig. 3. Effect of temperature and load on the linearized alternator and exciter time constants and gains

fications require that the system maintain at least ±0.5% accuracy from no load to 0.75 pf (power factor), lagging full load, and have as short recovery and settling times as possible over the temperature range −65 to +775 F.

The alternator must be capable of delivering 2-pu (per unit) 0.75-pf lagging load at 600 F external temperature which produces approximately 775 F in the control system. At this temperature it must also be capable of delivering 3-pu fault current into a zero-impedance balanced 3-phase fault. These conditions impose an alternator field current requirement which acts as a power requirement on the exciter. To obtain this exciter output power from a low-power-level error signal, two additional stages of amplification are required as shown in Fig. 2. With this configuration as a starting point, compensation networks must be designed to provide adequate system performance under changes of temperature and load.

For this particular program, a solid-rotor-design (with no brushes or rotating windings) 3-phase 400-cycle alternator, a similar rotating exciter, and magnetic amplifiers were used.

Because the exciter and alternator have copper windings whose resistance varies with temperature, the time constants and gains of the exciter and alternator vary by a factor of approximately 3½ under changes of temperature from −65 to 775 F. The time constant and gain of the alternator is further changed by a factor of approximately 2 due to load changes from no load to twice full load. The linearized effects of load and temperature on the cascaded exciter-alternator are shown in Fig. 3.

Examination of Fig. 3 shows that temperature and load changes cause more than a 20-to-1 gain change and shifts two time constants as much as 7 to 1 in the cascaded exciter-alternator transfer functions. To help offset these effects on system performance, a control signal proportional to alternator field current can be fed back to the interstage amplifier. This signal is returned to the interstage amplifier because of power level considerations. Use of I_{fd} for minor loop feedback compensation offers the advantage that this same signal can also be used in the torque-limiting technique discussed in a later section. Assuming that some form of series compensation will be used, the block diagram of the voltage regulator loop takes the form shown in Fig. 4.

Use of I_{fd} as a feedback signal causes the alternator gain change due to load changes to appear in K_L which is the incremental gain between alternator field current and alternator output voltage. The load change effect on the alternator time constant still remains in the minor loop. The revised temperature-load effects on the exciter-alternator combination are shown in Fig. 5.

To stay within the approximate component requirements that

$$1/500 \text{ second} < \text{time constants} < 1/2 \text{ second}$$

the feedback compensation network was selected as

$$H(s) = \frac{0.09(s/2.2+1)}{(s/500+1)}$$

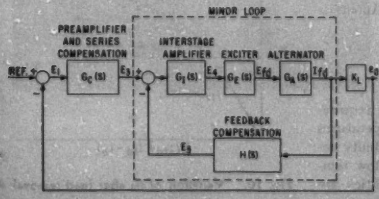


Fig. 4 (left). Block diagram of complete voltage regulator loop with alternator field-current feedback compensation

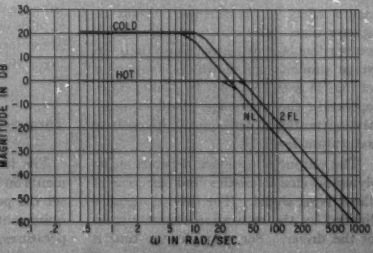


Fig. 5 (right). Effects of temperature and load on $I_{fd}/E_t = G_E(s)G_A(s)$ of Fig. 4

Fig. 6 shows the stability characteristics of the minor loop. It is common design practice to maintain stable minor loops so that in the event of major loop failure, the minor loop does not experience destructive behavior. Fig. 7 shows a plot of the minor loop forward and inverse feedback transfer characteristics which are used to obtain the approximate closed minor loop transfer characteristics shown in Fig. 8. The net effect of the feedback compensation is to reduce the effects of temperature and load changes that the major loop experiences and to shift these changes to regions that are fairly well above and below the desired range of major loop crossover frequencies.

To allow adequate system loop gain, a series compensation network (including the preamplifier) was selected as

$$G_c(s) = \frac{38.6(s/17+1)}{(s/2.2+1)(s/500+1)}$$

The over-all forward transfer characteristics are shown in Fig. 9. For changes in temperature from −65 to 775 F and load changes from no load to twice full load, the over-all regulating system has a phase margin from 40 to 60 degrees, a crossover frequency from 50 to 100 radians per second, and a loop gain in excess of 100.

Component Nonlinear Characteristics

The major types of nonlinearities encountered in this system are saturation and single-ended operation (a component produces an output for only one polarity of input signal). Alternator saturation provides a powerful means for limiting the output voltage rise after a heavy load or fault has been removed. In the steady-state loaded condition, $I_{kd}=0$ and I_{fd} is greater than no-load I_{fd} by an amount depending on i_q. At the time of load switching, i_q changes instantly and the total flux linkages remain constant, therefore, I_{fd} and I_{kd} experience negative steps or jumps. The I_{kd} jump decays rapidly to zero in the subtransient interval whereas

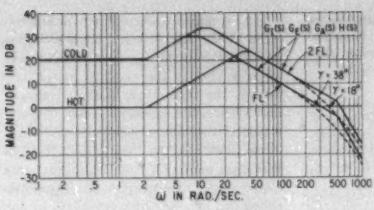


Fig. 6 (left). Minor loop stability analysis

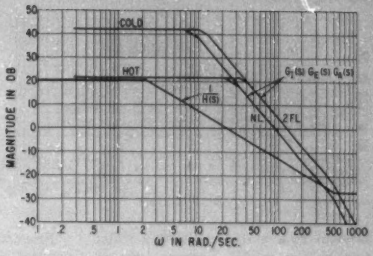


Fig. 7 (right). Minor loop transfer functions

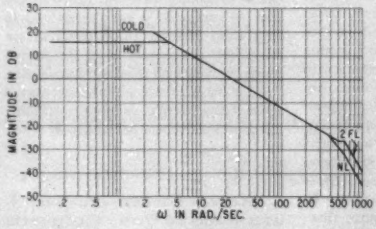


Fig. 8 (left). Closed minor loop transfer function

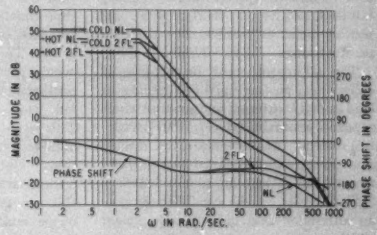


Fig. 9 (right). Bode diagram of entire forward transfer function

I_{fd} tends to rise to its previous value if the alternator field voltage is unchanged. The regulating loop takes action to reduce E_{fd} to the proper no-load value. Depending on the relative speed of response of the regulating loop and the field time constant, I_{fd} tends to rise and then fall to its steady-state no-load value. This variation in I_{fd} is shown in Fig. 10. As I_{fd} changes, Ψ_{md} changes in accordance with the saturation curve. If saturation is heavy, the change in Ψ_{md} for a given change in I_{fd} is less than when saturation is light. Since the output voltage varies with Ψ_{md}, saturation is very effective in limiting the output voltage rise during the transient interval. The same type of behavior is experienced upon load application except that the directions are reversed. This behavior can be seen in the computer runs of Figs. 11 and 12.

Because of size and weight considerations, the interstage amplifier and the rotating exciter are single-ended and voltage-limited. These nonlinearities have a major effect on lengthening system response time. The effects of single-ended versus double-ended operation of the interstage amplifier are shown by comparison of Fig. 11 with Fig. 12. In double-ended operation the interstage amplifier negative saturation is set equal to its positive saturation level.

The output of a device depends on its dynamic characteristics and the size of the driving signal. The driving signal is limited by the saturation characteristics of the driver. For example, the time it takes I_{fd} to reach a given value depends on the time constant and gain of the machine and the value of applied E_{fd}. When the system is hot, resistances are at a maximum and require highest voltages to produce required currents in the system. Therefore, voltage saturation in the interstage amplifier and exciter have the greatest effect on the system response at high temperatures and increasing voltage conditions. When the system is cold, the voltage levels to produce the required system currents are at a relatively low level. Therefore, the single-ended characteristics have a major effect under low temperatures and decreasing voltage conditions.

Single-ended operation and saturation produce the effect of a decrease in gain because a given input produces a smaller-than-linear output. The system tends to be slower for lower system gains. therefore these nonlinear effects increase the time for disturbances to settle out.

(The system block diagram and computer setup for this study are shown in Figs. 19 and 20, discussed in Appendix II.)

Torque-Limiting Techniques

The load- and fault-handling requirements on the system at high temperatures set the power and voltage capability requirements of the exciter. At low temperatures, these exciter capabilities are sufficient to produce high-current levels in the alternator. These high-current levels can produce excessive torques on the drive. Because signals proportional to developed torque are difficult to obtain, current level signals are used to limit developed torque. This portion of the study was concerned with the relative merits of alternator field current limiting versus load current limiting in effecting torque limiting. (The block diagram of the I_{fd} torque-limiting circuitry is shown in Fig. 19 of Appendix II. The i_a torque-limiting circuitry is similar except that i_a replaces I_{fd}.)

The system transient responses were obtained for application of unity pf faults. Fig. 13 shows a typical computer recording for application and removal of a fault. At the high temperature, exciter saturation is effective in limiting the tran-

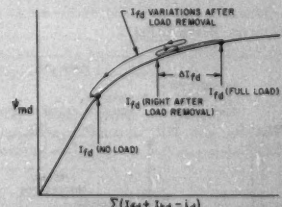


Fig. 10. Variation of I_{fd} after load removal

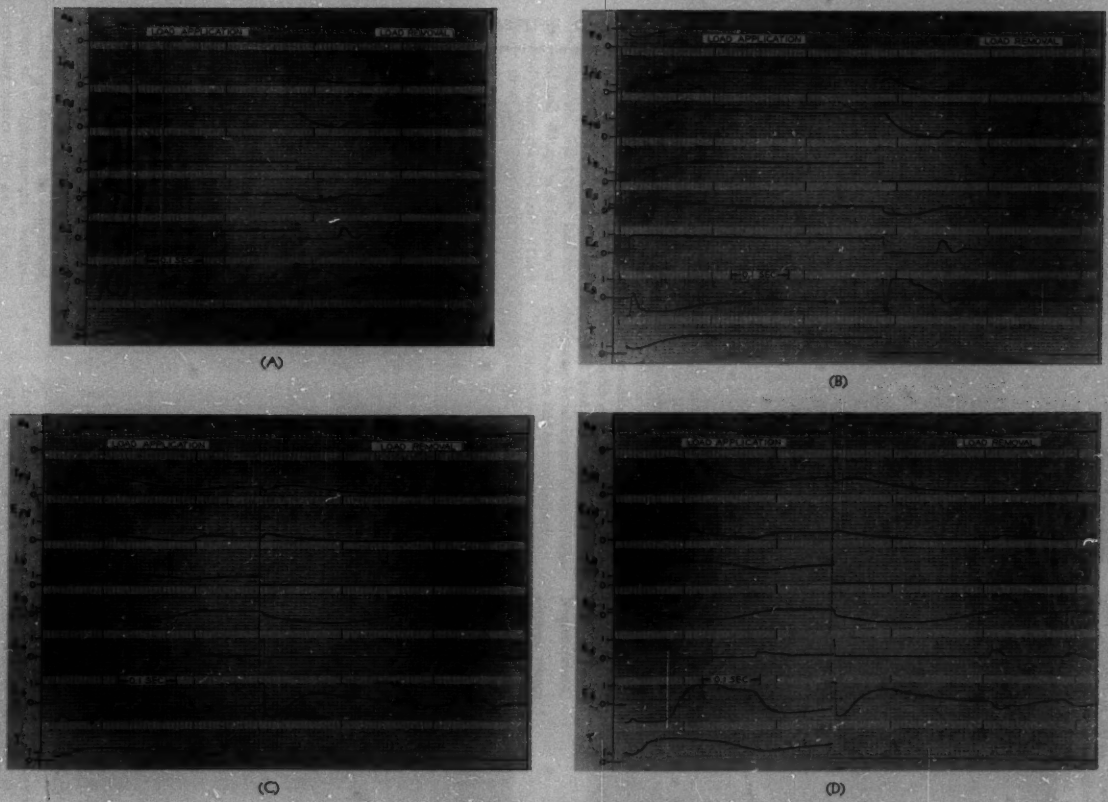


Fig. 11. System responses in pu for single-ended interstage amplifier

A—+775 F, 1 pu, 0.75-pf load B—+775 F, 2 pu, 0.75-pf load C—−65 F, 1 pu, 0.75-pf load D—−65 F, 2 pu, 0.75-pf load

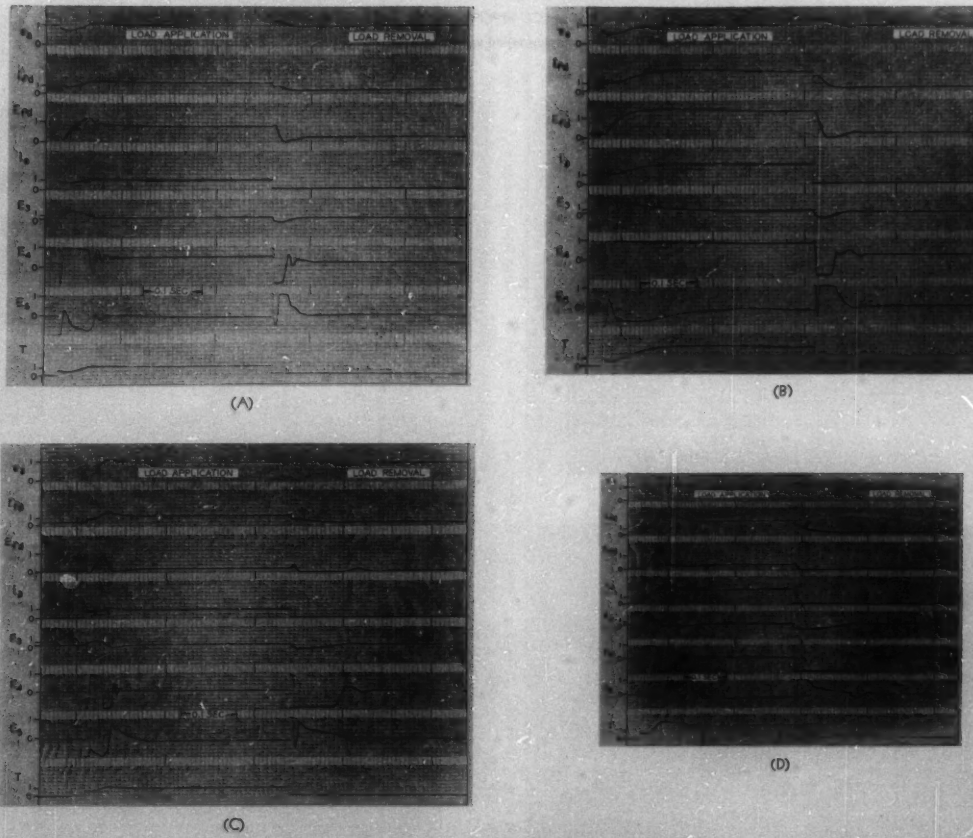


Fig. 12. System responses in pu for double-ended interstage amplifier

A—+775 F, 1 pu, 0.75-pf load B—+775 F, 2 pu, 0.75-pf load C——65 F, 1 pu, 0.75-pf load D——65 F, 2 pu, 0.75-pf load

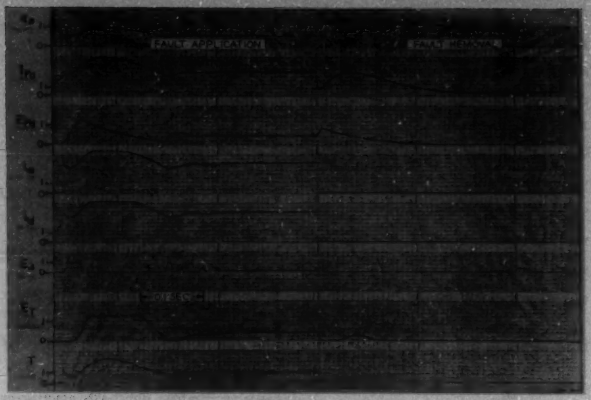


Fig. 13. System responses in pu at −65 F for 0.2-pu 1-pf balanced 3-phase fault application and removal

sient torque. From a steady-state viewpoint, the high-temperature condition is one of the worst cases because of high machine losses. From a transient viewpoint, the low-temperature condition is one of the worst cases because of the exciter capabilities. Figs. 14 and 15 show steady-state and peak torques, load current, and load voltage as a function of balanced 3-phase resistive fault. These figures show that I_{fd} limiting is more effective in limiting transient torque peaks. The torque peak occurs within 0.15 second after application of a fault and may or may not have serious consequences on the drive. Since I_{fd} is used in the regulating loop, this same signal measurement can be used to effect torque limiting.

Reactive Load Division Loop

At the time this paper was written, only a limited amount of work had been done with the reactive load division loop (power angle oscillation loop). A brief linearized analysis indicated that the possibility of power angle instability existed in the system. A detailed simulation approach for studying the reactive load division loop was developed and this approach is presented in Appendix IV.

Conclusions

This paper describes the analysis and design of a high-temperature voltage-regulating system. The study is based on the characteristics of a high-temperature alternator and control components at their present stages of development. The net effect of the high-temperature range is to restrict the range of component dynamic characteristics and to cause large changes in exciter and alternator transfer characteristics upon change of temperature and load. The effect of these changes can be minimized by use of negative alternator field current feedback.

The establishment of torque limiting by use of excitation system control shows promise. It is possible to reduce steady-state maximum torques to permissible values and to reduce maximum transient torques to short durations. However, evaluation against drive transient capabilities must be made before final selection of a method can be made. At present, for the two methods studied here, use of field current is preferable to use of armature current for control of maximum torques.

Simultaneous component nonlinearities were included in detail. Double-ended component operation can effect a major reduction in system response time upon application and removal of load.

A simulation approach was developed for study of the reactive load division loop.

In a voltage-regulating system, the disturbance is a step change in load. Since the characteristics of the power elements are known, it should be possible to apply nonlinear control design techniques to optimize the system response with respect to settling time and maximum deviation from rated voltage. Because of present component limitations, such optimization is impractical for high-temperature systems but should be considered for low-temperature systems and for high-temperature systems of the future. A promising new control component in this respect is the controlled rectifier. Its use should allow significant system accuracy and response improvements.

Appendix I

Nomenclature

e = phase voltage
i = phase current
r = phase resistance
I = field and amortisseur currents
R = field, amortisseur, and load resistance
a subscript = refers to armature
d subscript = refers to the direct axis
q subscript = refers to the quadrature axis
f subscript = refers to the field winding
k subscript = refers to amortisseur winding
o subscript = denotes base value
E_{fd} = field terminal voltage
X_{ad} = direct-axis mutual reactance
X_{aq} = quadrature-axis mutual reactance

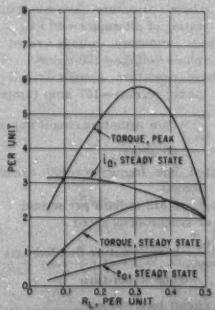


Fig. 14. Field-current torque-limiting curves

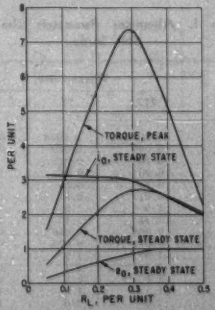


Fig. 15. Load-current torque-limiting curves

X_l = armature leakage reactance
X_f = field leakage reactance
X_{kd} = amortisseur direct-axis leakage reactance
X_{kq} = amortisseur quadrature-axis leakage reactance
X_d = stator direct-axis self-reactance
X_q = stator quadrature-axis self-reactance
$X_{f\!f}$ = field direct-axis self-reactance
X_{kkd} = amortisseur direct-axis self-reactance
X_{kkq} = amortisseur quadrature-axis self-reactance
X_{fkd} = mutual reactance between field and direct-axis amortisseur winding
X_d' = direct-axis transient reactance
$X_d = X_{ad} + X_l$
$X_q = X_{aq} + X_l$
$X_{f\!f} = X_{ad} + X_f$
$X_{kkd} = X_{ad} + X_{kd}$
$X_{kkq} = X_{aq} + X_{kq}$
$X_d' = X_{ad}X_f/X_{ad} + X_f + X_l$
X_L = load reactance
R_L = load resistance
Ψ_{md} = direct-axis mutual flux linkages
$\Psi_{f\!f}$ = total field flux linkages
Ψ_d = direct-axis flux linkages
Ψ_q = quadrature-axis flux linkages
Ψ_{kkd} = total direct-axis amortisseur flux linkages
ψ_{kkq} = total quadrature-axis amortisseur flux linkages
θ_L = load power-factor angle
δ = rotor power angle
ω = angular frequency
s = differential operator d/dt
t = time
T = air-gap torque
E = signal voltage
Q = reactive load kilovolt-amperes (kva)
T_D = drive torque kva
e_b = bus voltage

Normalized Synchronous Alternator Equations for Balanced 3-Phase Load

The following general alternator equations are used to obtain a detailed representation of the alternator.[4,5] All coefficients are in pu values. Field voltage equation:

$$E_{f\!f} = R_{f\!f} I_{f\!f} + s\Psi_{f\!f} \tag{1}$$

Amortisseur voltage equations:

$$0 = R_{kd} I_{kd} + s\Psi_{kd} \tag{2}$$

Table I. Alternator Parameters Used in This Study

	Temperature, F		
	775	77	−65
$R_{f\!f}$	0.0092	0.0038	0.0028
r	0.049	0.020	0.015
R_{kd}	0.15	0.062	0.045
R_{kq}	0.20	0.083	0.061
X_l	0.15	0.15	0.15
X_f	0.21	0.21	0.21
X_{ad}	1.9	1.9	1.9
X_{aq}	1.15	1.15	1.15
X_{kkd}	0.068	0.068	0.068
X_{kkq}	0.073	0.073	0.073
X_{fkd}	1.9	1.9	1.9
$X_{f\!f}$	2.1	2.1	2.1
X_d	2.1	2.1	2.1
X_q	1.3	1.3	1.3
X_{kkd}	2.0	2.0	2.0
X_{kkq}	1.2	1.2	1.2
X_d'	0.35	0.35	0.35

$$0 = R_{kq} I_{kq} + s\Psi_{kq} \tag{3}$$

Armature voltage equations:

$$e_d = s\Psi_d - \omega\Psi_q - r i_d \tag{4}$$

$$e_q = s\Psi_q + \omega\Psi_d - r i_q \tag{5}$$

Flux-linkage equations:

$$\Psi_{f\!f} = \Psi_{md} + X_f I_{f\!f} + (X_{fkd} - X_{ad}) I_{kd} \tag{6}$$

$$\Psi_{kd} = \Psi_{md} + X_{kd} I_{kd} + (X_{fkd} - X_{ad}) I_{f\!f} \tag{7}$$

$$\Psi_{kq} = (X_{aq} + X_{kq}) I_{kq} - X_{aq} i_q \tag{8}$$

$$\Psi_d = \Psi_{md} - X_l i_d \tag{9}$$

$$\Psi_q = X_{aq} I_{kq} - (X_{aq} + X_l) i_q \tag{10}$$

Ψ_{md} = nonlinear function of

$$[\Sigma(I_{f\!f} + I_{kd} - i_d)] \tag{11}$$

Load equations:

$$e_d = R_L i_d + \frac{X_L}{\omega_0} s i_d - X_L i_q \tag{12}$$

$$e_q = R_L i_q + \frac{X_L}{\omega_0} s i_q + X_L i_d \tag{13}$$

$$e_d^2 = e_d^2 + e_q^2 \tag{14}$$

$$i_a^2 = i_d^2 + i_q^2 \tag{15}$$

Torque equations:

$$T = i_q \Psi_d - i_d \Psi_q \tag{16}$$

$$Q = i_d \Psi_d + i_q \Psi_q \tag{17}$$

Per-Unit Quantities

To achieve the benefits of standardization, pu quantities are used in the machine equations. Per-unit quantities involve expressing the equation coefficients in terms of their respective bases. Unfortunately, there are several available basic approaches to defining pu values.[1,2,4] This situation effects a loss of some of the benefits of standardization.

For this study, the pu system is based on maintaining the same pu power in all windings, reciprocal mutual reactances, and use of peak values of voltage and current.

Armature circuit base values:

kva_{ao} = base armature kva = volt-ampere rating of alternator = 40 kva

e_{ao} = base armature voltage = rated peak phase voltage = 170 volts

i_{ao} = base armature current = rated peak phase current = 157 amp (amperes)

$$Z_{ao} = \frac{e_{ao}}{i_{ao}} = \text{base armature circuit impedance} = 1.08 \text{ ohms}$$

f_o = rated line frequency = 400 cycles per second

$\omega_0 = 2\pi f_o = 2{,}512$ radians per second

$\omega = 1$ at synchronous speed

$$\Psi_{ao} = \frac{e_{ao}}{\omega_0} = \text{base armature flux linkages} = 0.068 \text{ weber-turn}$$

Field circuit base values:

$\text{kva}_{f\!f o}$ = base field kva = kva_{ao}

$$I_{f\!f o} = \frac{i_{ao} X_{ad}}{X_{a\,f\!f}} = \text{base field current} = X_{ad} I_{f\!f} \text{ (no load)} = 4.75 \text{ amp}$$

$$E_{f\!f o} = \frac{\text{kva}_{f\!f o}}{I_{f\!f o}} = \text{base field voltage} = 8{,}420 \text{ volts}$$

$$Z_{f\!f o} = \frac{E_{f\!f o}}{I_{f\!f o}} = \text{base field impedance} = 1{,}770 \text{ ohms}$$

$$\Psi_{f\!f o} = \frac{E_{f\!f o}}{\omega_0} = \text{base field flux linkages} = 3.37 \text{ weber-turns}$$

Direct-axis amortisseur circuit base values:

kva_{kdo} = base direct-axis amortisseur kva = kva_{ao}

$$I_{kdo} = \frac{i_{ao} X_{ad}}{X_{akd}} = \text{base direct-axis amortisseur current}$$

$$E_{kdo} = \frac{\text{kva}_{kdo}}{I_{kdo}} = \text{base direct-axis amortisseur voltage}$$

$$Z_{kdo} = \frac{E_{kdo}}{I_{kdo}} = \text{base direct-axis amortisseur impedance}$$

$$\Psi_{kdo} = \frac{E_{kdo}}{\omega_0} = \text{base direct-axis amortisseur flux linkages}$$

Quadrature-axis amortisseur circuit base values:

kva_{kqo} = base quadrature-axis amortisseur kva = kva_{ao}

$$I_{kqo} = \frac{i_{ao} X_{aq}}{X_{akq}} = \text{base quadrature-axis amortisseur current}$$

$$E_{kqo} = \frac{\text{kva}_{kqo}}{I_{kqo}} = \text{base quadrature-axis amortisseur voltage}$$

$$Z_{kqo} = \frac{E_{kqo}}{I_{kqo}} = \text{base quadrature-axis amortisseur impedance}$$

$$\psi_{kqo} = \frac{E_{kqo}}{\omega_0} = \text{base quadrature-axis amortisseur flux linkages}$$

The coefficient values in Table I are in pu. The machine saturation curve is shown in Fig. 16. Rated no-load armature voltage without machine saturation requires a field current of 2.5 amp.

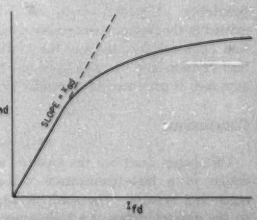


Fig. 16. Machine no-load saturation curve

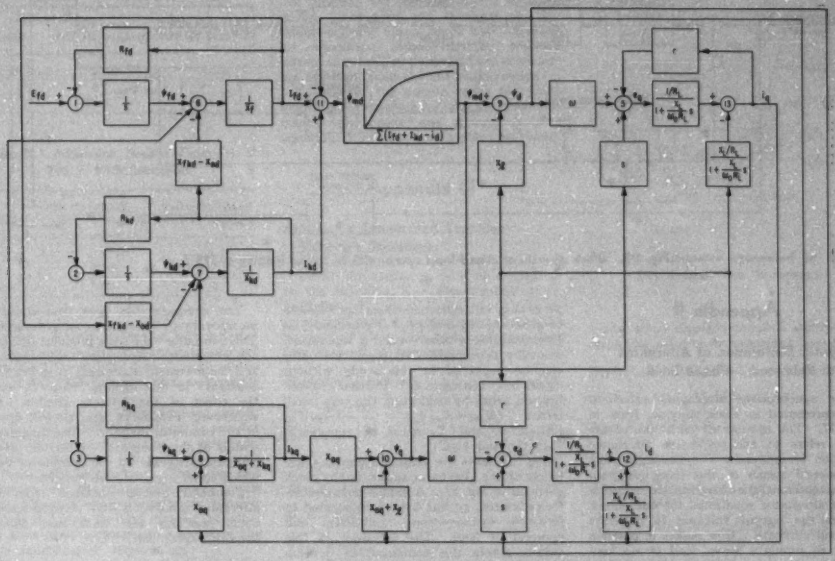


Fig. 17. Complete synchronous alternator block diagram with balanced 3-phase load

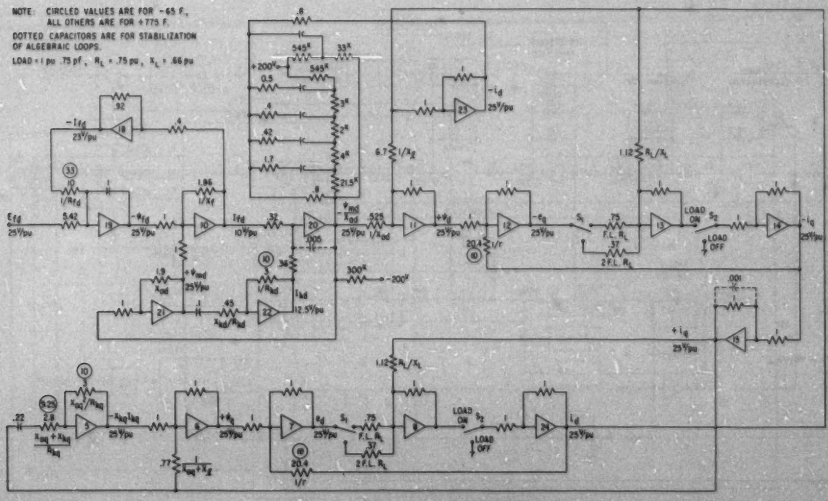

NOTE: CIRCLED VALUES ARE FOR −65 F., ALL OTHERS ARE FOR +775 F.
DOTTED CAPACITORS ARE FOR STABILIZATION OF ALGEBRAIC LOOPS.
LOAD = 1 pu .75 pf, R_L = .75 pu, X_L = .66 pu
LOAD ON
LOAD OFF
F.L. R_L
2 F.L. R_L

Fig. 18. Computer diagram of alternator

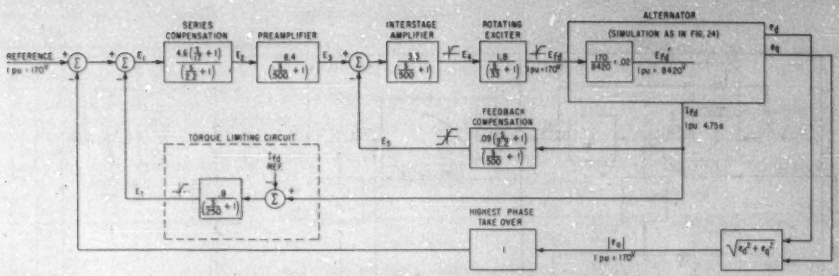


Fig. 19. Block diagram of closed-loop system with I_{fd} torque limiting at 775 F

Appendix II

Computer Simulation of Alternator with Balanced 3-Phase Load

The synchronous alternator equations are represented in block diagram form in Fig. 17. The number at each summation point refers to the respective alternator equation in Appendix I. Although there are several forms of the block diagram which represent the simultaneous solution of the alternator equations, the one shown utilizes the output voltages to generate the load currents. This makes it possible easily to simulate heavy load or no load on the alternator (corresponding to an R_L change from a very small value to infinity).

A disadvantage of representing the alternator on the analog computer in the form shown by the block diagram of Fig. 17 is that two differentiations are required to generate $s\Psi_d$ and $s\Psi_q$. Fortunately, as recommended in other work,[4] it was experimentally determined that these two terms can be neglected in this study without significant consequence. Further simplification arises by neglecting the very small terms $(X_{fad}-X_{ad})I_{fd}$, $(X_{fd}-X_{ad})I_{fd}$, $(X_L/\omega_0)\,si_d$, and $(X_L/\omega_0)\,si_q$ in equations 6, 7, 11(B), and 11(C).

Fig. 18 shows the analog computer simulation of the alternator shown in the block diagram of Fig. 17. A double-pole double-throw switch, s_1, has been incorporated to facilitate instantaneous application and removal of load. The inclusion of this switch allows the computer to develop its own initial conditions, either with or without load, after which the load may be changed and the resulting transient observed. Switch s_1 provides a convenient method of changing the load which is to be applied and removed.

The computer has been time-scaled to incorporate a 250-to-1 time scale expansion. This computer time scale provides sufficient spread between the system time constants and the necessary stabilizing capacitor time constants in the algebraic loops to make the effect of the latter negligible. The stabilizing capacitors are shown dotted in the computer diagram. The magnitude scaling of the computer is indicated on the computer diagram by the number of volts which equal 1 pu for each variable.

Computer results obtained from the alternator simulation were checked against actual machine test data and showed excellent agreement.

Simulation of Voltage Regulator System with Torque Limiting

Fig. 19 shows the block diagram of the closed-loop system as designed earlier in this paper, and Fig. 20 above the computer

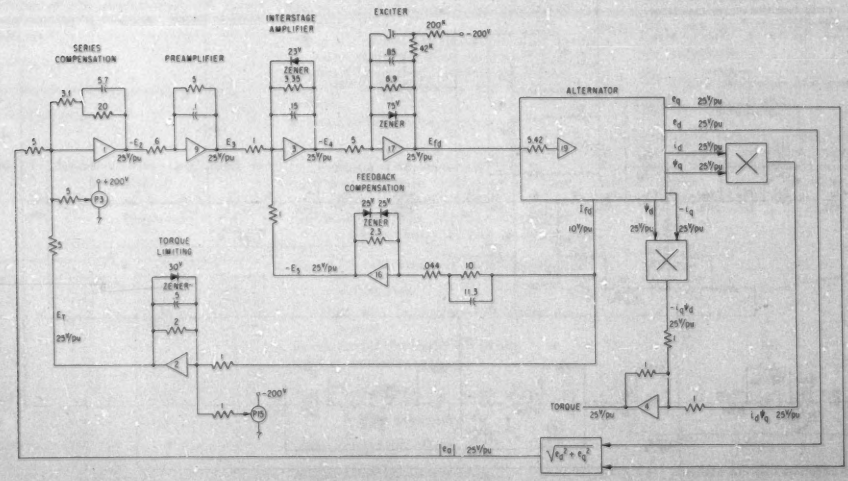


Fig. 20. Computer diagram of closed-loop system shown in Fig. 19

Table II. Alternator Transfer Function Without Saturation

No Load	Full Load (0.75 PF Lagging)	Twice Full Load (0.75 PF Lagging)
$\frac{4.16}{\left(\frac{s}{11}+1\right)}$	$\frac{1.39}{\left(\frac{s}{25.3}+1\right)}$	$\frac{0.84}{\left(\frac{s}{35.8}+1\right)}$

Table III. Alternator Transfer Functions at 775 F With Saturation

No Load	Full Load (0.75 PF Lagging)	Twice Full Load (0.75 PF Lagging)
$\frac{1.68}{\left(\frac{s}{22}+1\right)}$	$\frac{1.01}{\left(\frac{s}{30}+1\right)}$	$\frac{0.79}{\left(\frac{s}{42}+1\right)}$

diagram. A scale factor has been included at the input to the alternator to convert from the 170-volt base used throughout the loop to the 8,420-volt base of the alternator field.

The output of the torque-limiting circuit, E_T, is single-ended and is therefore zero for values of I_{fd} less than the I_{fd} reference. Under normal load operation this circuit is essentially inoperative. The gain and reference of the torque-limiting circuit were adjusted experimentally to provide maximum torque limiting without introducing limit cycle oscillations or violating the 3-pu fault-current requirement.

The simulation of the closed-loop system with I_{fd} torque limiting is shown in Fig. 18. The time and magnitude scaling are the same as in Fig. 18. Torque limiting was also studied using i_a feedback instead of I_{fd}. The i_a is generated by obtaining $\sqrt{i_d^2+i_q^2}$ with a function generator and this signal is fed into amplifier 2 in place of I_{fd}. The gain and reference of this torque-limiting circuit were also adjusted for maximum torque limiting without introducing limit cycle oscillations or violating the 3-pu fault-current requirement.

To observe the torque which the drive would experience under various transient conditions, torque was generated from equation 16 using the multipliers indicated.

Appendix III

Alternator Linearized Transfer Function Equations

Under the assumptions of no saturation in the machine, a constant-speed drive, negligible $s\Psi_d$, $s\Psi_q$, $(X_L/\omega_0)si_d$, $(X_L/\omega_0)si_q$ terms, and negligible amortisseur effects, the normalized alternator and load equations (see Appendix I) become:

$$E_{fd}=R_{fd}I_{fd}+s\Psi_{fd} \tag{18}$$

$$e_a=-\Psi_q-ri_d \tag{19}$$

$$e_q=\Psi_d-ri_q \tag{20}$$

$$\Psi_{fd}=X_{fd}I_{fd}-X_{ad}i_d \tag{21}$$

$$\Psi_d=X_{ad}I_{fd}-X_di_d \tag{22}$$

$$\Psi_q=-X_qi_q \tag{23}$$

$$\Psi_{md}=X_{ad}(I_{fd}-i_d) \tag{24}$$

$$e_d=R_Li_d-X_Li_q \tag{25}$$

$$e_q=R_Li_d+X_Li_q \tag{26}$$

$$e_a^2=e_d^2+e_q^2 \tag{27}$$

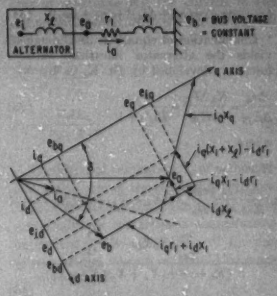


Fig. 21. Alternator connected to infinite bus through line impedance

For small changes about a steady-state operating point these equations become:

$$\Delta e_d=X_q\Delta i_q-r\Delta i_d \tag{28}$$

$$\Delta e_q=X_{ad}\Delta I_{fd}-X_d\Delta i_d-r\Delta i_q \tag{29}$$

$$\Delta e_d=R_L\Delta i_d-X_L\Delta i_q \tag{30}$$

$$\Delta e_q=R_L\Delta i_q+X_L\Delta i_d \tag{31}$$

$$\Delta E_{fd}=\frac{R_{fd}}{X_{ad}}\left\{s[X_{ad}\Delta I_{fd}-(X_d-X_d')\Delta i_d]\times\frac{X_{fd}}{R_{fd}}+X_{ad}\Delta I_{fd}\right\} \tag{32}$$

$$e_{at}\Delta e_a=e_{dt}\Delta e_d+e_{qt}\Delta e_q \tag{33}$$

Equation 33 is obtained by expanding equation 34 and neglecting the second-order effects.

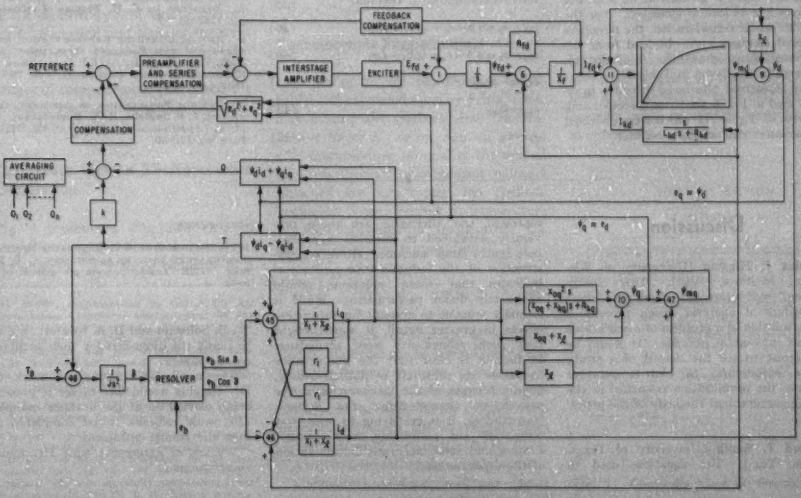


Fig. 22. Reactive load division simulation block diagram

$$(e_{ds}+\Delta e_d)^2=(e_{ds}+\Delta e_d)^2+(e_{qs}+\Delta e_q)^2 \quad (34)$$

Combination of equations 28 through 33 yields the alternator linearized transfer function (developed by Dr. K. G. Black):

$$\frac{\Delta e_s}{\Delta E_{fd}}=\frac{K_L\Delta I_{fd}}{\Delta E_{fd}}=\frac{K_1}{T_{do}'s+1} \quad (35)$$

where

$$K_1=\frac{\dfrac{e_{do}X_{ad}}{E_{fdo}R_{fd}}\left\{\dfrac{e_{qs}}{e_{ds}}[X_L(X_L+X_q)+R_L(R_L+r)]\times\dfrac{e_{ds}}{e_{qs}}[X_qR_L-X_LR]\right\}}{(R_L+r)^2+(X_L+X_q)(X_L+X_d)} \quad (36)$$

$$T_{do}'=\left[\frac{(R_L+r)^2+(X_L+X_d')(X_L+X_q)}{(R_L+r)^2+(X_L+X_d)(X_L+X_q)}\right]T_{do}' \quad (37)$$

$$T_{do}'=\frac{X_{fd}}{\omega_oR_{fd}} \quad (38)$$

Here e_{ds}, e_{ds}, and e_{qs} are the pu steady-state values calculable from the steady-state vector diagram. The alternator parameters are in pu values obtained from Table I. Equation 35 yields the alternator transfer function whose gain and time constant have physical dimensions because in equation 36 the ratio of the base values, e_{do}/E_{fdo}, is included and in equation 38, ω_o is introduced.

At +775 F, use of equations 35 through 38 yields the alternator transfer functions shown in Table II.

Saturation has a major effect on the machine characteristics in that it reduces the machine gain and time constant. The saturation effect is greatest at no load and becomes less as load is increased. For the machine under consideration, the responses to small changes were obtained from the computer setup shown in Appendix II. The results are shown in Table III.

The linearized alternator transfer functions used in Fig. 3 are those corresponding to those of Table III which were obtained by measurement on the computer.

Appendix IV

Reactive Load Division Loop Simulation

When operating alternators in parallel, the electrical system must insure proper division of reactive load between machines whereas the drive system must insure proper division of real load between machines. Alternator characteristics are such that while delivering a given average kva to the load, a sizable reactive kva can oscillate between alternators while the bus voltage remains invariant at its normal level. The reactive load division loop must provide stabilizing signals which insure proper division of reactive load among alternators without reactive load oscillations.

The following analysis (developed by F. P. deMello) provides a means for simulating the reactive load division loop during parallel operation of alternators.

In parallel operation, machines tend to pair off and exchange load. The behavior of one machine is the mirror image of its partner and the bus voltage remains invariant. A load stability study can therefore be made by considering one machine against an infinite bus which is located at the electrical center between machines.

The circuit diagram in Fig. 21 shows the situation where the lines between the alternator and bus have an impedance inscrted for purposes of generality. The vector diagram for this situation is also shown in Fig. 21. For purposes of simpler presentation, the assumption is made that the small alternator internal resistance, r, is negligible or that it can be lumped with the line resistance to produce r_1. Here e_f is the alternator air-gap voltage while X_1 is the line reactance. The equations for this configuration become:

$$e_{bd}=e_b\sin\delta=e_d+i_qX_1-i_dr_1 \quad (39)$$

$$e_{bq}=e_b\cos\delta=e_q-i_dX_1-i_qr_1 \quad (40)$$

$$e_{id}\simeq-\Psi_{mq} \quad (41)$$

$$e_{iq}\simeq\Psi_{md} \quad (42)$$

$$e_f\simeq\Psi_d \quad (43)$$

$$e_d\simeq-\Psi_q \quad (44)$$

$$e_{bd}=-\Psi_{mq}+(X_1+X_l)i_q-i_dr_1 \quad (45)$$

$$e_{bq}=\Psi_{md}-(X_1+X_l)i_d-i_qr_1 \quad (46)$$

$$\Psi_{mq}=\Psi_q+X_li_q \quad (47)$$

$$\delta=\frac{1}{Js^2}(T_D-T) \quad (48)$$

These equations, in conjunction with equations 1, 6, 9, 10, and 11 of Fig. 17 and the regulating loop diagram of Fig. 4, allow generation of the block diagram shown in Fig. 22. In this diagram, T_D can be inserted as a constant or as a time-varying function, depending on the drive characteristics. The reactive load signal, Q, is compared with the average reactive load supplied to the bus per machine. The difference signal is used to control the reactive load that the alternator supplies to the bus. The compensation network and a portion of the real load signal, T, are used for stabilization purposes. At the time of writing this paper, this simulation had not been verified by test.

References

1. Synchronous Machines—Theory and Performance (book). C. Concordia. John Wiley & Sons, Inc., New York, N. Y., 1951.

2. Parallel Operation of Airplane Alternators. D. W. Exner. *AIEE Transactions (Electrical Engineering)*, vol. 62, Dec. 1943, pp. 755–60.

3. A Basic Analysis of Synchronous Machines—Part I. W. A. Lewis. *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 77, Aug. 1958, pp. 436–56.

4. Analog Computer Representations of Synchronous Generators in Voltage-Regulation Studies. M. Riaz. *Ibid.*, vol. 75, Dec. 1956, pp. 1178–84.

5. *Discussion* by C. H. Thomas of reference 4, pp. 1182–84.

6. Analog Computer Representation of Alternators for Parallel Operations. L. V. Boffi, V. B. Haas, Jr. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, May 1957, pp. 153–58.

7. Random Paralleling of Aircraft Alternators. F. P. deMello, H. M. Rustebakke. *Ibid.*, pt. II (*Applications and Industry*), vol. 77, Nov. 1958, pp. 322–30.

Discussion

Thomas J. Higgins (University of Wisconsin, Madison, Wis.): This clearly written, well-integrated paper evidences the values of analytic design procedures in the solution of a problem of considerable current interest in practice. It would be of interest to see the details of a study which encompass, for low-temperature systems, the optimization remarked in the closing paragraph of the body of the paper.

Richard T. Smith (University of Texas, Austin, Tex.): The equations used to represent the alternator are those previously developed to describe a salient-pole synchronous machine. However, construction of the inductor-type alternator[1] indicates that these equations should adequately define performance. While it is easily possible to account for saturation effects in greater detail, it was thought that such refinements were premature. In fact, it is clear that the performance equations are presently of limited utility, mainly because of the inaccuracies of the parameters characterizing the system. Nevertheless, it is gratifying to note that computers are permitting, economically, a more and more accurate determination of these parameters.

Reference

1. A Solid Rotor A-C Generator for High-Temperature Electrical Systems. J. T. Bateman. *AIEE Transactions*, see pp. 400–05 of this issue.

W. E. Sollecito and D. A. Swann: We wish to thank the discussers for their comments on our paper.

The nonlinear optimization work that Dr. Higgins would like to see is presently being carried on at the authors' company. The work remains to be completed and then the results published.

We are in agreement with Dr. Smith's observations.

An Elementary Design Discussion of Thermoelectric Generation

E. W. BOLLMEIER
ASSOCIATE MEMBER AIEE

AT THE PRESENT TIME, many electric power sources of unconventional nature are being pursued vigorously in research and development. One of these, thermoelectric generation, has enjoyed renewed interest because of the development of semiconductor thermoelectric materials yielding efficiencies which are now practical for some devices and encouraging for others.

Semiconductor thermoelectric materials of practical use today have multiplied thermoelectric heat conversion efficiencies by a factor of 10, approximately, compared with metal thermocouples. This remarkable increase has brought thermoelectric conversion to a point of potential competition with electromechanical or other conversion for some applications. However, large-scale economic competition remains dependent on significantly higher materials efficiency attainment. At today's state of the art, the uses to which thermoelectric generating devices are suited are those which value some feature more highly than maximum economy or which utilize very low-cost heat.

This paper seeks to acquaint the reader with the elementary problems of thermoelectric generation and some of the solutions devised. Consideration is also given to the advantages and disadvantages of thermoelectric generation for uses which may appear desirable at the present state of the art.

A Composite View of Thermoelectric Generation

Fig. 1 presents a simplified view of thermoelectric generation as a complete, useful system. The usual elements of a useful heat energy conversion system are apparent.

1. Heat input.
2. Conversion apparatus.
3. Heat rejection.
4. Output control.

These same four elements of a useful system may be readily compared to the 1. burner, 2. boiler-turbine-generator, 3. condenser, and 4. control system of a conventional steam generating system. Similar problems and similar general solutions apply equally to both, although the elements are physically different.

The major difference between conventional electromechanical generation and thermoelectric generation is that the boiler-turbine-generator conversion devices are replaced by one unit, the thermopile.

The thermopile, being a static device, avoids many mechanical problems inherently. However, substantial new problems appear as the struggle for realization of the semiconductor materials efficiency in practical generators is carried out.

Thermocouple Fundamentals

According to the relationships established and expanded upon by Fritts, Ioffe, and others, the Seebeck emf (electromotive force) E_{sc} or open-circuit voltage of a thermocouple is equal to the sum of the two thermocouple element voltages obtained by integrating the Seebeck coefficient S with respect to temperature between hot-junction temperature T_h, and cold-junction temperature T_c:[1,2]

$$E_{sc}=\int_{T_c}^{T_h} S_p dT-\int_{T_c}^{T_h} S_n dT \qquad (1)$$

The Seebeck coefficients of semiconductor materials are in general quite temperature-dependent. For accuracy in calculation of E_{sc}, the integral in equation 1 must be used, but for simplification of the expressions to follow, average values of Seebeck coefficient multiplied by (T_h-T_c) will be used rather than the integral.

Connecting a thermopile of open-circuit voltage V_g, and resistance R_g, into a circuit of resistance R_L, the current in the circuit and power output may be defined by:

$$I=\frac{V_g}{(R_g+R_L)} \qquad (2)$$

$$P_o=\frac{V_g^2 R_L}{(R_g+R_L)^2} \qquad (3)$$

The internal resistance R_g of a thermopile is the sum of the element resistances. Like Seebeck coefficient, the resistivity of semiconductor materials is temperature-dependent. The calculation of thermopile resistance, therefore, should be made by the integral

$$R_g=N\left[\frac{1}{A_p}\int_{T_c}^{T_h}\rho_p\frac{dx}{dT}dT+\frac{1}{A}\int_{T_c}^{T_h}\rho_n\frac{dx}{dT}dT\right] \qquad (4)$$

where

N = number of couples
A = element area
ρ = electrical resistivity

Equation 4 can be evaluated when the variation of resistivity with temperature and temperature distribution within the thermocouple material are known. The following expression which defines the relation of R_g to R_L at the condition of maximum efficiency:

$$R_g=\frac{R_L}{\sqrt{1+Z\frac{(T_h+T_c)}{2}}} \qquad (5)$$

where Z, figure of merit

$$=\frac{(S_p-S_n)^2}{(\sqrt{k_n\rho_n}+\sqrt{k_p\rho_p})^2} \qquad (6)$$

where Z is available plotted as a function of temperature for both P and N materials, the average Z value over the temperature range encountered may be used in equation 5. Z_{avg} is determined by:

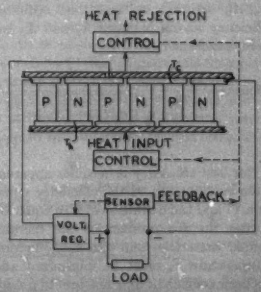


Fig. 1. Composite view of thermoelectric generation

Paper 59-867, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 24, 1959; made available for printing April 22, 1959.

E. W. BOLLMEIER is with the Minnesota Mining and Manufacturing Company, St. Paul, Minn.

The author wishes to express his appreciation to Mr. E. J. Levin, Mr. W. G. Krawczak, and Mr. R. A. Elm for their contributions to the contemporary state of the art in thermoelectric generation. He also gratefully acknowledges the orientation provided by Dr. R. W. Fritts in successful use of the products of his semi-metals research.

$$Z_{avg}=\frac{\int_{T_c}^{T_h} Z_p dT+\int_{T_c}^{T_h} Z_n dT}{2(T_h-T_c)} \quad (7)$$

It may also be shown for maximum efficiency that the cross-sectional areas of the thermocouple elements are related as follows when $L_n=L_p$, which is usually desirable in thermopile design:

$$\frac{A_n}{A_p}=\sqrt{\frac{k_p\rho_n}{k_n\rho_p}} \quad (8)$$

With A_n/A_p established and thermocouple resistance $R_e=R_g/N$ known for any proposed thermopile design, approximation of the length to area ratios for the P and N thermocouple legs may be made through the relations:

$$\frac{L_n}{A_n}=\frac{R_g/N}{\rho_n+\frac{(\rho_p A_n)}{A_p}} \quad \text{When } L_n=L_p \quad (9)$$

$$\frac{L_p}{A_p}=\frac{R_g/N}{\rho_p+\frac{(\rho_n A_p)}{A_n}} \quad \text{When } L_n=L_p \quad (10)$$

The physical size and weight of the thermocouple elements can be varied without affecting thermopile electrical design characteristics, so long as the ratios of L/A are observed. The final choice of exact L and A dimensions may be made on the basis of mechanical feasibility and available heat flux.

Thermocouple operating efficiency is expressed as the ratio of useful power delivered, I^2R_L, to the rate of heat input to the thermocouple. Rate of heat input is made up of loss rate by thermal conduction from hot junction to cold junction, input rate to drive the thermoelectric cycle at the Carnot efficiency established by T_h and T_c and half the I^2R loss of the thermocouple material circuit. The expression for maximum conversion efficiency is given by the following relation, subject to the conditions that optimum A_n/A_p ratio exists and that optimum efficiency match between R_g and R_L has been established:

$$\eta_E(\max)=\frac{T_h-T_c}{T_h}\,\frac{\sqrt{1+Z(T_h+T_c)/2}-1}{\sqrt{1+Z(T_h+T_c)/2}+\frac{T_c}{T_h}} \quad (11)$$

For best efficiency it becomes apparent from examination of the efficiency equation 11 and the definition of Z in equation 6 that the $K\rho$ product of the thermocouple materials, and the cold-junction temperature T_c must be minimized and Seebeck coefficient and hot-junction temperature T_h maximized. Since for most electrical conductors and semiconductors, reduction in thermal conductivity is accompanied by approximately proportional increase in electrical resistivity, reducing the $k\rho$ product while maintaining or increasing S is a difficult materials research problem.

Today's practical semiconductor thermoelectric materials, such as the p- and n-type lead tellurides, have $k\rho$ products ranging up to 10 times those of metals. Fortunately, conversion efficiency improves with the square of Seebeck coefficient and inversely with the first power of the $k\rho$ product. Thus, the tenfold advantage in semiconductor Seebeck coefficient over those of metals yields a tenfold improvement in efficiency over metals, in spite of the higher semiconductor $k\rho$ products. Further reference to Fritts[1] is recommended for his broad consideration of p- and n-type lead telluride thermocouple design.

Thermopile Configuration

Physical arrangement of a thermopile structure is primarily dependent upon:

1. Heat-source energy density and configuration.
2. Radiator absorption capacity and configuration.
3. Heat losses and surface transfer rates.
4. Mechanical and electrical design feasibility.

Heat-source energy density has inherent bearing upon thermopile configuration because of the direct relation between energy density and size of heat source. The heat source must be designed to the thermopile's heat input requirements and should be of the minimum area necessary to accommodate the number of elements of the required L/A ratio in a mechanically and thermodynamically feasible arrangement. It is fundamental that heat insulation losses will be minimized if a maximum of the heat conducting path from source to radiator is made up of thermoelectric material. Available heat-source energy density and transfer rates to the thermopile hot junctions determine whether or not this ideal can be approached.

Assuming complete coverage of the heat source with thermoelectric material, configuration and type of heat source determine the degree of insulation heat loss to be combated. If the heat source were in the form of a thin slab, it might be covered with close-packed parallel elements having a very short heat-transfer path between elements from hot to cold junction. If the heat source were long and cylindrical, it might be covered with radially arrayed elements which present a small area of nonelement heat conducting path near the source but increasing with radial distance. A third heat-source configuration is the sphere, surrounded by elements extending to a spherical radiator.

The configuration choice of the thermopile design engineer is thus controlled by the factors of feasible heat-source configuration; energy density of source and minimum heat-leakage path.

Considering the spherical configuration concept, and a form of heat energy of given energy density per unit volume, it becomes apparent that a spherical heat-source configuration will contain maximum energy per unit of surface area compared to cylindrical or slab-source configurations. In heat sources where the rate of energy liberation per unit volume is low, a problem of mismatch between high source area and low thermoelectric material area may occur. Any source configuration other than spherical will present greater surface area for the same rate of energy liberation and same energy density.

The insulation heat loss of the spherical concept thermopile is proportional to the mean area A_m of the insulation body. A_m is characterized by the following relation, assuming negligible heat transfer from elements to insulation along their length:

$$A_m=\sqrt{(A_s-A_{tm})(A_E-A_{tm})}$$
$$=\sqrt{(\pi D_s^2-A_{tm})[\pi(D_s+2t_i)^2-A_{tm}]} \quad (12)$$

where

A_s = source area
A_E = insulation exterior area
A_{tm} = total cross-sectional area thermoelectric material, assuming constant element diameters
t_i = insulation thickness

In the range of $t_i=D_s$ and $A_{tm}=D_s^2/2$, the A_m relation simplifies to:

$$A_m=\pi\frac{\sqrt{17}}{2}D_s^2$$

Thus, it is shown for the spherical configuration that insulation heat loss is proportional to the square of heat source diameter in a one-dimensional ratio area of mechanical feasibility.

Similar relationships can be shown for other thermopile configurations, but it may be shown that insulation heat loss for all configurations will vary exponentially greater than unity with heat-source size.

Heat-Source Discussion

In the case of combustible fuel sources—whether or not coupled with

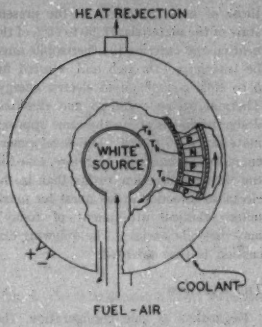


Fig. 2. Incandescent heat source, spherical configuration

heat-transfer mediums—uniform hot-junction temperature, combustion efficiency, and flue losses present major problems which can be attacked in four ways:

1. High energy density incandescent combustion system.
2. High-temperature closed-cycle condensing-vapor boiler system.
3. Flue gas to fuel, air heat exchanging.
4. Nuclear-energy heat sources.

The high energy density incandescent combustion system (Fig. 2) is one which seeks to transfer energy to the hot junction by radiation in preference to convection and conduction. Theoretical consideration of this system shows advantages from the standpoints of combustion efficiency, maximum energy density (minimum size), and low thermal inertia. Incandescent sources designed to transfer heat appreciably by radiation offer advantage also in designing for uniformity of hot-junction temperature, a task extremely difficult with gas convection or conduction transfer.

High-temperature condensing-vapor boiler systems (Fig. 3) offer advantages of hot-junction temperature uniformity and high heat-transfer rates at lower temperatures than required by fluid conduction or convection. This may be important in protecting thermoelectric materials which could be damaged easily by accidental overfiring of a direct flame heated thermopile. Temperature uniformity can be obtained by heating the pile with condensing vapor while energy is stripped from the heat source over a fair range of temperature approaching boiler temperature. The problems of closed-cycle high-temperature boilers are primarily those of containing the most probable candidates, such as mercury, protecting against overpressure, and minimizing thermal inertia for control purposes.

In regard to heat exchanging between flue gas and incoming fuel-air mixtures, greater advantage can be gained if flue gas stack draft can be used to provide the pressure necessary to the exchanger. Since electric power is available at rather low efficiency from the pile, use of electric output for powering inefficient accessories may yield insignificant improvement in over-all efficiency, particularly in small output generators.

The nuclear heat source is one of inherent practicality for thermoelectric generation. Nuclear sources, deriving energy from radioactive decay (in the case of isotopes) or fission (in the case of reactors), have no problem of flue gas loss to hamper over-all efficiency. The nuclear heat source has exceptional watt-hours of energy content per unit weight, compared to all other heat sources—an advantage for long life generator service in remote areas, in space, or in locations where air is not available for combustion. The nuclear source inherently minimizes a number of problems for the thermopile design engineer. Because it needs no physical connection with the pile exterior for flue gas discharge and fuel inlet, heat conducting members other than thermoelectric material or thermal insulation can be essentially eliminated. The disadvantages of nuclear sources for many ordinary uses today are cost and availability. Yet they are indispensable for those uses demanding maximum watt-hours per pound or where periodic refueling is impossible.

Much remains to be known regarding long-term radiation effects on semiconductor thermopiles. Insufficient knowledge is available to state effects on all semiconductors though it has been predicted by Pritts and others that crystal lattice damage should annal out rapidly at the thermopile operating temperatures involved.

The SNAP III nuclear-fueled generator reported on by Anderson, Barmat, and the author operated without measurable deterioration of thermoelectric properties since fueled on January 7, 1959, to the present. No great significance can be drawn from this except to note that the source external radiation level was low, very little neutron emission was involved and no damage of the doped lead telluride elements was expected, nor has it occurred.[3] Significant studies of radiation damage to n and p doped lead telluride materials are now being conducted by a number of firms active in nuclear technology.

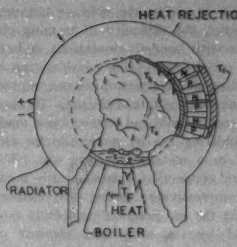


Fig. 3. Closed boiler heat source

Designing with Semiconductor Thermoelectric Materials

Prime semiconductor material factors influencing thermopile design are:

1. Stability and strength.
2. Contact resistance.
3. Sublimation.
4. Oxidation and contamination.

Designing with materials at elevated temperatures involves compromises resulting from the temperature dependency of most material properties. The same problems apply to semiconductor thermoelectric materials in general, though the discussion following is based on the result of experience with p and n doped lead tellurides.

Semiconductors of thermoelectric value are, in general of brittle nature and poor in tension, but of good strength in compression. Table I gives the physical properties data of p and n doped lead telluride.

The strength properties of these materials suggest that the pile structure should be so designed as to limit thermocouple stresses to compression. This compromise imposes limitations, but at the same time aids in the solution of another problem, that of contact resistance.

Contact resistance may be minimized for doped lead telluride materials through the employment of bonded or soldered connection at the cold end and pressure contact at the hot junction. Elements may be formed with integrally bonded contacts

Table I. Typical Physical Properties, Cast and H_2 Annealed Doped Lead Telluride

Tensile strength, pounds per square inch	> 1,000
Compressive strength, pounds per square inch	> 10,000
Young's modulus, pounds per square inch	2×10^6
Density, grams per cubic centimeter	8.15
Thermal expansion coefficient, per degree C	18×10^{-6}
Thermal conductivity, per centimeter per degree C	0.02 watts

for use at either the cold or hot end. However, the difficulty of forming elements with bonded connections at both ends makes engineering consideration of an all-bonded connection scheme impractical at today's state of the art. In usual practice, the cold-end connection is made by soldering of the semiconductor element to a metal contacting cap. The hot-end connection is satisfactorily made by compressively loading the element against the hot-junction connecting shoe shared by its companion element of opposite polarity. Constructions utilizing bonded hot-junction contacts are possible, but unnecessary except where pressure loading is undesirable or inconvenient. Flat-module pile constructions may face mechanical problems in compressive element loading because of the bending strength of the hot and cold-end plates. In such designs, bonded hot-junction contacts and soldered cold contacts may be exploited to some advantage. Atmospheric pressure contact loading of an evacuated flat pile module may also be exploited, but contacting pressure is limited to values considered insufficient at today's state of the art.

Pressure-loaded elements assure that electrical continuity will be maintained even though some sublimation may occur at the highest temperature portion of the elements. Sublimation may be regarded as an operating life design problem at temperatures in excess of 1,000 F (degrees Farenheit), and although it may be minimized by providing the elements with a coating or cover that restricts diffusion of material away from the hot junction, pressure contacting assures maximum reliability.

The factors of oxidation and poisoning of doped lead tellurides are also temperature-dependent in large degree. Both p and n lead tellurides can be used in extended air exposed service up to 500 F. However, beyond that point oxidation of the material becomes significant and hermetic sealing is required. Since operating temperatures above 500 F are desirable for maximum Carnot efficiency and the material is capable of stable operation in excess of 1,000 F with long life, hermetic sealing is indicated for maximum over-all efficiency designs.

Oxidation above 500 F may be eliminated as a problem in thermopile design by providing a reducing, inert, or high-vacuum atmosphere within the hermetically sealed enclosure. A reducing atmosphere protects both the element material and all electrical contact surfaces against oxidative degradation. While high vacuum may also provide oxidation protection, it also raises the curve of sublimation rate versus temperature. The increased structural, mechanical, and material problems created by designs which seek to maintain high vacuum over a long life are difficult to solve and costly. High-vacuum environment does have the advantage of lowering losses through most thermal insulation materials which might be used in pile design. However, over-all efficiency gain through use of vacuum is complex and questionable after consideration of the problems of increased sublimation.

Contamination of most thermoelectric materials at elevated temperatures is a problem of which the thermopile designer must be constantly aware. Contamination may result from the diffusion of a number of metals, such as zinc and copper, into the material or from reaction with a large group of active elements and radicals. Diffusion of contaminants into the thermoelectric material results predominately in lowered Seebeck coefficient, whereas reactive contamination causes increased contact resistance or sublimation rate. The best guide for safety in design is experimental testing of contamination by materials apt to come in intimate contact with the thermoelectric material at elevated temperatures.

In general, hydrogen or methane is desirable as an atmospheric environment for doped lead telluride above 500 F. Such reducing atmospheres serve to protect against oxidation or reactive contaminants that may diffuse in gaseous state to the materials. Materials such as iron, molybdenum, alumina, carbon, and mica are safe for direct contact at elevated temperature. Aluminum, tin, copper, and steel are satisfactory construction materials at the cold end of the pile where solid or gaseous diffusion proceeds at negligible rates. Thermal insulations should in general be regarded as suspect until it has been experimentally determined that they are inert at elevated temperatures. Insulation surface coatings or binders may also cause contamination.

Thermopile Structure Design

Until thermoelectric materials significantly beyond the present 8 to 11% efficiency range are available, the thermoelectric device engineer's lot will be one of limited ability to compromise. His first problem today is tolerable output energy cost and his second is one of heat transfer—too much and too little.

As material conversion efficiencies improve, thermopile design problems decline at a considerably higher rate, particularly those of heat transfer. At the present state of the materials art, 90 to 95% of the heat energy entering the thermopile must be transferred through and rejected for 5 to 10% conversion to electric energy. Thermal design problems and electrical design problems are in the same approximate ratio: 90 to 10. The development engineer's ingenuity is, therefore, taxed in the thermal direction, rather than in the electrical direction in his quest for maximum practical attainment of today's materials efficiencies. The following discussion is thus oriented.

The Hot End

Regardless of pile configuration, the hot-end problem, is one of minimum impedance to heat transfer from source wall to thermoelectric element, in order that source wall temperatures may be kept below safe levels for practical materials of construction. In addition, the element must be electrically insulated from the source (if conductive) and its neighbors, but joined electrically to its companion element of opposite polarity. Solution to the simultaneous problems of low impedance to heat flow and reliable electrical insulation of the interface between element hot-junction connecting shoe and heat-source wall appears best found in the area of high-density vitrified ceramic materials. Heat-flow impedance at this interface is the sum of the interface and insulation thermal resistances. The electrical insulation thermal resistance, proportional to thermal conductivity and thickness, is reduced by use of thin vitreous coatings on the element hot-junction connecting shoes. Also, by use of such bonded coatings, one mechanical interface between shoe and source wall is eliminated.

The pressure contact between element and hot junction shoe is a lesser problem of thermal impedance not only because it is a single interface, but also because sufficient plastic flow does take place in lead telluride thermoelectric material at temperatures over 900 F to produce partial, mechanically bonded contact to the hot-junction shoe. Reducing atmospheres encourage this bonding by reducing at operating temperature the oxides that may be present on the interface materials. This reduction of interface oxides by reducing atmosphere may be observed in the drop of internal resistance observed in the first few hours of operation of a pressure-contacted doped lead telluride thermopile. It has been noted that the thermal impedances of all the mechanical interfaces between heat source and radiator are lowered by internal gas pressure and raised with vacuum. Gas molecules

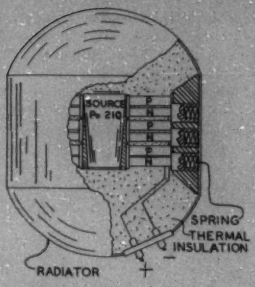


Fig. 4. SNAP III generator, 3M-1 thermopile

aid the transfer of heat across dry mechanical interfaces, thus lowering interface themal impedance.

Between the Ends

As previously considered in the section on configuration, thermal insulation losses can be best limited by covering the source wall area with a maximum of thermoelectric material. Where a match can be designed between thermal watts needed per unit area of element and thermal watts available per unit area, this ideal can be approached according to the ingenuity of the designer in shaping elements for close packing and arranging hot junction connecting shoes. Usually, the 1-to-1 area match ideal cannot be realized for reasons of configuration, mechanical design feasibility, or cold-end transfer capacity.

The Cold End

Cold-end thermal impedance, like that of the hot end must be minimized while reliable electrical insulation is maintained between elements and radiator wall. Fig. 4 illustrates one scheme of cold-end design for pressure-contacted elements.

The design of Fig. 4 utilizes a bonded metallic cold-end element contact. This contact member is in mechanical contact with a pressure applying member which is electrically insulated from the radiator hardware by means of a thin coating. Other schemes of greater or fewer mechanical interfaces may be readily devised depending on configuration, radiator wall area available per element, necessity of hermetic sealing and cost. The required features of any pressure-contacted element cold-end design are 1. reliable electrical insulation and 2. mechanical flexibility in the thermal path from element to radiator wall sufficient to permit individual element differential movement under spring loading.

The design of Fig. 4 aids electrical insulation reliability by avoiding compressive loading of the electrical insulation. It provides mechanical flexibility in the cold junction thermal path through the sliding fit between the pressure-applying member and the radiator wall. Thermal impedance of this mechanical interface is reduced by its relatively large area and a grease-like coating applied to the interface at assembly. Other schemes involving flexible metallic bellows, diaphragms, etc., may have advantages dependent upon obtainable areas, thicknesses and thermal conductivities of the interfaces, electrical insulation, and hardware involved.

Heat transfer from the radiator wall to the available heat sink is a complex subject which may be attacked in a number of ways with or without liquid or gaseous transfer mediums. Since the transfer rate of all heat-transfer methods—convection, conduction and radiation—is proportional to area, it is advantageous to design for maximum radiator area consistent with other problems. Generator designs for use in space environment face limitations in radiator area and weight. These limitations raise the temperature driving force necessary for the required radiation heat dump rate, with resultant lowering of generator Carnot efficiency. The spherical and cylindrical generator configurations have the inherent advantage of higher external area than flat configurations for the same area of thermoelectric material. The thermopile designer may find it often to his advantage to utilize cylindrical or spherical configuration to ease the problem of cold-end heat transfer, and at the same time benefit from the structural strength, minimum weight and reduced difficulty of hermetically sealing "round" configurations.

Output and Control

The problem of output control is essentially one of heat input and radiator rejection rate control. The ability to control thermal input varies with type of heat source from moderate for direct fired to not-at-all for nuclear isotope sources. Less variation exists in ability to control heat rejection rates. Also the problem of thermal inertia in thermoelectric generation at the present state of the art makes output response to load changes sluggish.

Inertia can be reduced by shrinking element size while holding L/A constant for any generator design, provided that source energy density and all heat transfer rates can be increased proportionately. These provisions are increasingly difficult as size of element goes down. Where fast response to load change is needed, the best solution may be found in the use of an electric accumulator between generator and load or by control switching of a reserve portion of the pile to match the load. Like most of the designer's problems, control response will be aided by the reduction of weight which will be made possible by higher efficiency materials.

The difficulty of rapid response control of thermoelectric generators typical of today's state of the art is such that they should be considered best suited to stable or slowly varying loads, except where battery accumulators can be used for instantly available reserve power.

Because of the fact that thermocouples are inherently low-voltage high-current devices, the problem of designing to high-voltage low-current loads becomes difficult. Since high-voltage low-current output requirements lead to high generator internal impedance design, the number of thermoelectric elements goes up and their L/A ratio becomes high. As the L/A ratio goes up, elements become long and thin, thus creating problems of mechanical strength for the designer and making element fabrication difficult. As design power goes up, the problem resolves itself since most electric power is utilized at under 200 volts, regardless of amperage. It may be stated as a generalization that generator designs for output characteristics of less than 10 watts and more than 10 volts would be simplified, if static d-c to d-c voltage conversion were used to achieve design output voltage. Prospective users of thermoelectric generation at low power levels should consider the voltage problem in their system design and strive for voltage-to-current

Table II. Operational Characteristics of the 3M-1 Thermopile

	Initial*	At Source Half-Life
Thermal input, watts	95	48
Fueling, curies	3,000	1,500
Dose rate, generator surface, millirontgens per hour	700	350
Dose rate, at 1 meter, milliroentgens per hour	7	3.5
Hot junction temperature, degrees F	1,100	720
Cold junction temperature, degrees F	400	170
Electrical output, watts	5.3	2.4
Voltage at maximum efficiency	3.0	2.4
Over-all efficiency, %	5.5	5.1
Watt-hours, first half-life	= 5,000	
Watt-hours/pound, first half life = 1,500		

* Estimated. Actual fuel load was initially at half-life values.

ratios of less than unity if possible in the 10- to 50-watt ranges.

The SNAP III Generator

The SNAP III nuclear-fueled generator, Fig. 4, is a working example of practical thermoelectric generation.[3,4] It utilizes a type *3M-1* thermopile of cylindrical concept built so as to receive a tapered cylindrical isotope (polonium 210) container as its heat source. Since the SNAP III was designed to demonstrate feasibility of thermoelectric power supply for space vehicles, heat rejection is by radiation only. The design of this thermoelectric generator established many of the concepts outlined here. Its operational characteristics are given in Table II. The thermoelectric element and thermopile specifications are as follows:

THERMOELECTRIC ELEMENT SPECIFICATIONS

Material: p and n doped lead telluride
A_n/A_p ratio: 0.89
L, element length: 1.0 inch
R_{pn}, resistance per couple: 0.07526 ohm
E_{ac}, couple voltage: 190 millivolts
Segmentation: n yes, p no

THERMOPILE SPECIFICATIONS

T_h, temperature, hot junction: 1,100 F
T_c, temperature, cold junction: 400 F
P_o, power output: 5 watts
V_p, voltage output, maximum efficiency: 2.8 volts d-c
R_i, internal resistance: 2.03 ohms
R_L, load resistance, maximum efficiency: 2.12 ohms
Number of couples: 27
Heat source: polonium 210, 138-day half-life
Diameter: 4.75 inches
Height: 5.5 inches
Weight: 5 pounds

This generator is working proof of practical thermoelectric power. It should be noted that the efficiency attained was the result of a combination of good efficiency thermoelectric materials, substantial design and a 100% thermally efficient heat source. The demonstrated efficiency may be debated from the standpoint of economics, but it remains a tribute to those who brought about its existence. This same group has since built and operated an advanced design generator of greater than 7% over-all efficiency.

Conclusions

The problems confronting the development and design of thermoelectric generators which extract maximum usefulness from today's semimetal thermoelectric materials have been presented. The discussion of these problems serves only to define their probable boundaries. The solutions suggested indicate only that the problems have been attacked, not solved.

An example has been given as proof that the concepts defined herein have been reduced to practice and that thermoelectric generation is practical today for many applications which are difficult or uneconomical to accomplish with conventional power-generation methods.

References

1. DESIGN PARAMETERS FOR OPTIMIZING THE EFFICIENCY OF THERMOELECTRIC GENERATORS UTILIZING P-TYPE AND N-TYPE LEAD TELLURIDE, R. W. Fritts. *AIEE Transactions*, pt. I (Communication and Electronics), vol. 78, Nov. 1959, pp. 817–20.

2. THERMOELECTRIC GENERATORS, A. F. Ioffe. *Izvest Akadamii Nauk SSSR*, Moscow, USSR, no. 1, Jan. 1956.

3. SNAP III—PRACTICAL POWER FROM RADIOISOTOPES, M. Barmat, G. M. Anderson, E. W. Bellmeier. *Nucleonics*, New York, N. Y., vol. 17 no. 5, 1959.

4. STATUS REPORT ON THERMOELECTRICITY, H. F. Straus, B. B. Rosenbaum, P. Maycock. *Memorandum Report 901*, Naval Research Laboratory, Washington, D. C., Feb. 1959, p. 21.

New Materials for Transformer-Rectifier Unit in High-Temperature Aircraft

J. G. HOOP
MEMBER AIEE

D. K. McILVAINE
MEMBER AIEE

AIRCRAFT operating at multi-Mach speeds and high altitudes subject some portions of the electric systems to severe environmental changes. A program is underway to develop a suitable system including an a-c generator with constant-speed drive, a contactor, a circuit breaker, a current transformer, and a transformer-rectifier unit for conversion to direct current. Development and selection of materials for use in the transformer-rectifier are covered in this paper.

Paper 59-883, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 24, 1959; made available for printing April 23, 1959.

J. G. HOOP and D. K. MCILVAINE are with the Westinghouse Electric Corporation, Lima, Ohio.

This work was done at Materials Engineering, Research Laboratory, and Aircraft Equipment Department of Westinghouse Electric Corporation. Air Force Contract No. *AF33(600)35489* with North American Aviation.

Environment

Electric equipment operated at a 315 C (degrees centigrade) ambient temperature in the low-density air at 80,000 feet altitude may develop hot spots several hundred degrees above the ambient. Storage of aircraft at temperatures far below freezing subjects the electric equipment to moisture condensation and produces pronounced changes in the electrical and mechanical properties of the materials. The materials to be used in this program are evaluated and chosen on the basis of their individual properties and on their compatibility with the other materials. In general, only inorganic materials can be expected to have any appreciable life at the higher temperatures to be encountered.

Performance

The transformer-rectifier unit being developed for high-temperature service is to convert power from a 200-volt 3-phase 400-cycle supply to give a 28-volt d-c output at 50-ampere load. A radio noise filter network is to be provided. Electrical performance of this unregulated lightweight transformer-rectifier unit restricts the minimum num-

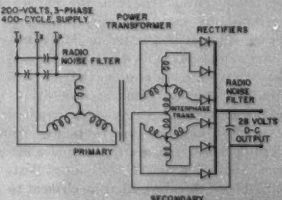


Fig. 1. Circuit of 50-ampere transformer-rectifier power supply

Fig. 2. Test core assembly

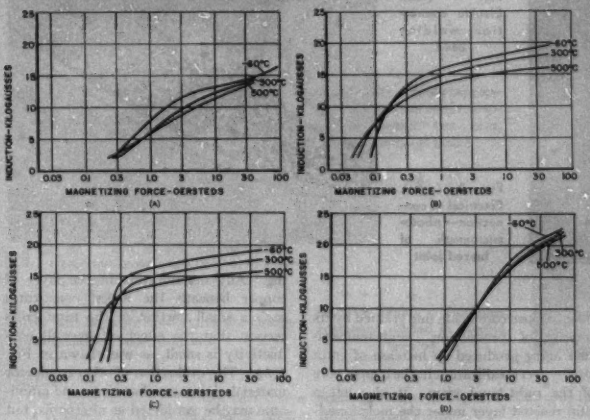


Fig. 3. Magnetization curves

A—4.7% silicon unoriented iron 0.014-inch thick
B—3% silicon oriented iron 0.012-inch thick
C—3% silicon oriented iron 0.004-inch thick
D—27% cobalt oriented iron 0.025-inch thick

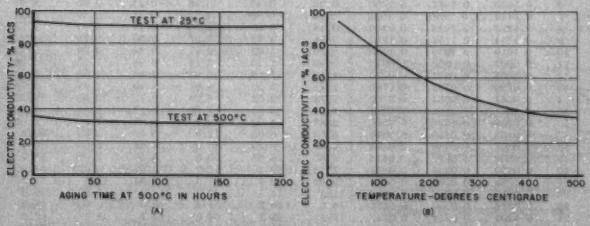


Fig. 4. Conductivity tests of no. 17 AWG copper wire with 0.006-inch nickel plate

A—Versus aging time at 500 C
B—Versus temperature

ber of phases to six. Mechanical considerations discourage the use of a higher number. Regulation and efficiency requirements eliminate consideration of bridge-type circuits with high-temperature rectifiers because of the double voltage drop thus introduced.

Circuit

Fig. 1 shows a 6-phase double-Y connection with an interphase transformer which permits two rectifiers to operate in parallel at all times, thus reducing the rectifier loading and transformer secondary kilovolt-ampere rating. Although delta-primary connections are preferred for rectifier applications on the basis of a-c harmonic considerations, the Y connection was chosen for this relatively low-power unit because it maintains the output voltage better under a single-phase fault condition. With only two secondary windings per phase, the coil build-up is held to a minimum, and interleaved windings are easily made. Consequently, a low leakage reactance is obtained. Radio noise filtering, as shown, is achieved by straight capacitance filter networks across the input and output circuits. This form was chosen to obtain minimum weight and complexity.

Magnetic Materials

Choice of a core material from the presently available ferromagnetic alloys called for further investigation of their properties over the broad temperature range to be encountered in service. Special core assemblies suitable for use in an oven were prepared by the steps shown in Fig. 2. Magnetization curves at 500 C, 300 C, and −60 C are given in semilogarithmic plots for 4.7% silicon unoriented iron in Fig. 3(A), 3% silicon oriented iron in Fig. 3(B), (C) for heavy and thin gages respectively, and 27% cobalt oriented iron in Fig. 3(D). Even under heavy magnetizing forces the 4.7% silicon unoriented iron achieves only moderate induction. The 3% silicon oriented iron in both the heavy and thin gages attains a higher maximum induction, but it is quite temperature sensitive except in the region where the curves cross at an induction of about 10 kilogausses. Curves for the 27% cobalt oriented iron show it to be the least temperature sensitive and capable of higher maximum induction than the others. The fact that a relatively high magnetizing force is required with this latter material is not a serious disadvantage for application in this program because the

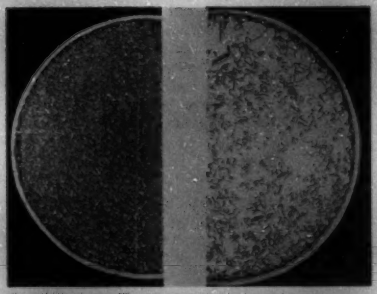


Fig. 5 (left). Effect of temperature aging on nickel-plated copper

Fig. 6 (right). Effect of temperature aging on conductor joints

Fig. 7 (left). Electric connection—welding bead

Fig. 8 (right). Electrical connection—photomicrograph of brazed joint

higher induction contributes to weight reduction. However, it should be noted that if a nuclear environment were present, a material with cobalt should not be used because it becomes activated.

Wire

Because the expected service life at high temperature is only 20 hours, a high degree of oxidation resistance is not required in the magnet wire to be employed in the transformer windings. Tests indicate that copper wire is adequately protected when covered by a 0.0006-inch plating of nickel. The relatively minor effect of 500 C aging in air on the electric conductivity of this type of wire is shown in Fig. 4(A). The bottom curve in that figure and the one in Fig. 4(B) serve to illustrate the fact that copper, like all other common conductors, has a temperature coefficient which causes its conductivity to decrease to less than half of its room temperature value when heated to 300 C. Photomicrographs from etched cross sections of nickel-plated copper wires before and after aging for 200 hours at 500 C give a visual picture in Fig. 5 of the metallurgical effects. In the left-hand view, the outer layer of nickel covers an inner copper core made fine grained by its process of production. High-temperature aging produced an increase of grain size by recrystallization in the copper core of the right-hand view. It also left a thin reacted layer under the nickel coating. Any deterioration in the layer of copper beneath the nickel constitutes such a small portion of the total cross section that the effect on over-all conductivity is small, as was shown in Fig. 4(A). The evaluation of other conductor materials indicates that silver and aluminum may be considered as alternates, but

Table I. Magnet Wire

Copper, 0.0006-Inch Nickel Plate, 400 C Insulation

Wire Gage	Wire Diameter			Bare Copper Resistance Ohms/M-ft.			Plated Wire Resistance Ohms/M-ft.		
	Bare	Plated	Insulated	−55 C	+20 C	400 C	−55 C	+20 C	400 C
6	0.1620	0.1633	0.1700	0.28	0.40	0.98	0.28	0.39	0.98
7	0.1443	0.1456	0.1520	0.35	0.50	1.24	0.35	0.49	1.24
8	0.1285	0.1298	0.1360	0.44	0.63	1.56	0.44	0.62	1.55
9	0.1144	0.1157	0.1216	0.56	0.79	1.97	0.56	0.79	1.96
10	0.1019	0.1032	0.1090	0.70	1.00	2.49	0.69	0.99	2.48
11	0.0907	0.0920	0.0975	0.89	1.26	3.14	0.88	1.25	3.12
12	0.0808	0.0821	0.0873	1.12	1.58	3.96	1.11	1.55	3.93
13	0.0720	0.0733	0.0783	1.41	2.00	4.98	1.40	1.98	4.79
14	0.0641	0.0654	0.0702	1.78	2.52	6.29	1.77	2.50	6.23
15	0.0571	0.0584	0.0631	2.24	3.18	7.92	2.22	3.16	7.70
16	0.0508	0.0521	0.0566	2.83	4.02	10.0	2.80	3.98	9.90
17	0.0453	0.0466	0.0510	3.56	5.05	12.6	3.52	4.98	12.5
18	0.0403	0.0416	0.0458	4.50	6.39	15.9	4.44	6.30	15.7
19	0.0359	0.0372	0.0413	5.67	8.05	20.0	5.59	7.93	19.7
20	0.0320	0.0333	0.0371	7.12	10.1	25.2	7.02	9.93	24.8
21	0.0285	0.0298	0.0335	9.02	12.8	31.9	8.84	12.5	31.3
22	0.0253	0.0266	0.0303	11.4	16.2	40.3	11.2	15.9	39.5
23	0.0226	0.0239	0.0273	14.3	20.3	50.6	14.0	19.6	49.2
24	0.0201	0.0214	0.0247	18.2	25.7	64.0	17.7	25.1	62.8
25	0.0179	0.0192	0.0224	22.8	33.4	80.8	22.2	31.5	78.4
26	0.0159	0.0172	0.0203	28.9	41.0	102	27.9	39.6	98.6
27	0.0142	0.0155	0.0183	36.2	51.4	128	34.9	49.6	123
28	0.0126	0.0139	0.0166	46.0	65.3	163	44.2	62.7	156
29	0.0113	0.0126	0.0152	57.2	81.2	202	54.7	77.4	194
30	0.0100	0.0113	0.0138	73.2	104	260	69.6	98.6	246

Fig. 9 (left). Coating tower for magnet wire

Fig. 10 (right). Vacuum impregnating furnace for insulating compounds

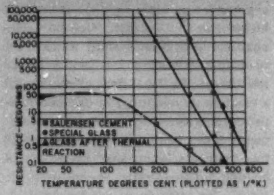


Fig. 11 (left). Insulating compounds resistivity versus temperature; potted test coils, vacuum impregnated

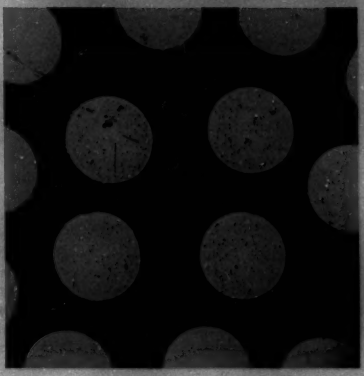


Fig. 12 (right). Insulating compound reacted in electric winding

silver will be more costly and aluminum connections are more difficult.

Joints

A variety of techniques have been tried for producing twisted conductor joints as shown in Fig. 6. Row (A) illustrates the condition of electric joints before test and row (B) the change caused by aging 100 hours at 540 C. These specimens were prepared by 1. arc welding in inert gas, 2. brazing with 72% silver and 28% copper, 3. brazing with 50% silver, 16% zinc, 18% cadmium, and 16% copper, 4. brazing with 50% silver, 34% copper, and 16% zinc, 5. brazing with 80% copper, 15% silver, and 5% phosphorus, and 6. brazing silver with 80% copper, 15% silver, and 5% phosphorus. The welded junctions deteriorated very little as a result of aging. A photomicrograph in Fig. 7 shows the general good condition of the weld bead and its junction with the conductor which was protected by the nickel coating extending intact into the bead. The brazing alloys containing phosphorus deteriorated badly, but the high silver alloys produced a good joint. A photomicrograph from a section of the fourth joint, as shown in Fig. 8, is typical of the better brazed joints.

Wire Insulation

Successful wire insulation for coating of nickel-plated copper must produce a suitable adherent insulating layer on the final conductor in order to be capable of withstanding deformation associated with the winding process and also retain good properties when heated in the finished winding. The desired combination of properties has been attained by the use of an inorganic pigment in a resin coating on the magnet wire. The first four columns in Table I give the nominal diameters of "bare," "plated," and "insulated" wires of this type for a range of sizes from 6–30 (American Wire Gage).

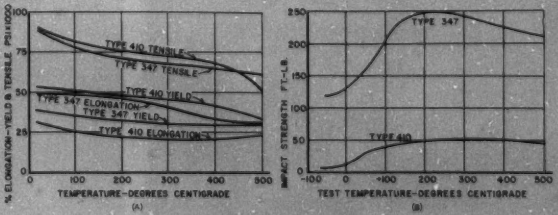


Fig. 13. Structural materials

A—Mechanical properties versus temperature

B—Impact strength versus temperature for stainless steel

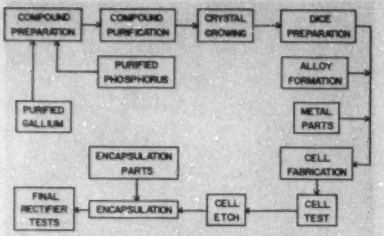


Fig. 14. Semiconductor crystals development flow chart for gallium phosphide rectifiers

Fig. 15 (right). Crystal pulling furnace

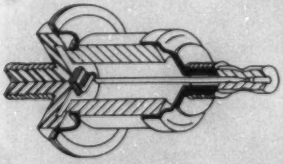


Fig. 16. High-temperature rectifier device configuration

In the remaining columns the tabulation shows a comparison of bare and nickel-plated wire resistances at temperatures of −55 C, +20 C, and 400 C. It may be noted that the plated wire has an appreciably lower resistance than the corresponding bare wire in the higher gages. This is due to the fact that the same nominal gage size has been here assigned to wires with the same copper cross section. The change in resistance between −55 C and +400 C is more than threefold. Application of this type of wire insulation has been done in an experimental enameling tower whose bottom or entering end appears in Fig. 9. Heating a winding of wire so coated breaks down and drives off a portion of the resin to leave a deposit forming an inorganic coating on the wire.

Subsequent vacuum impregnation with glass in a furnace of the type shown in Fig. 10 produces a solid winding structure impervious to penetration by moisture. Impregnation involves a preliminary heating under vacuum in the upper furnace section, a dip into molten glass, and retraction into the upper hot zone where admission of inert gas or air forces molten glass into the winding's voids. Simple glasses formed of low melting oxides penetrate best during impregnation because they are more fluid than those with higher melting points. However, the lower melting point glasses generally have lower volume resistivity at a given temperature than do those with higher melting points. Special glass mixtures have been devised which react upon further heating and yield a higher resistivity. In Fig. 11 the effect of such a reaction on resistivity at elevated temperatures is shown by circles for measurements taken before and by triangles for measurements taken after the extended heating. The increase in resistivity at 400 C, for example, is one-hundredfold. The curve represented by squares gives the resistivity values for a common commercial cement, as a comparison. Wires embedded in the reacted compound are shown in Fig. 12.

Structural Materials

Structural members need to maintain good mechanical properties over the full range of temperatures to be encoun-

Fig. 17. Vacuum furnace for coating metal foil

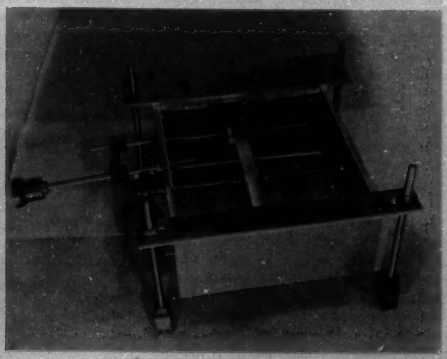


Fig. 18. Foil reeling fixture

Fig. 19 (left). Foil winder for condenser assembly

Fig. 20 (right). Capacitor encapsulations

tered. The elongation, yield, and tensile strengths of two typical alloy steels are shown in Fig. 13(A). Impact strengths of the same alloys, austenitic type *347* and martensitic type *410*, are given in Fig. 13(B). The advantage of the type *347* alloy steel is particularly evident in its superior impact strength at low temperatures.

Rectifiers

Semiconductors to be considered for use as high-temperature rectifiers need to be capable of operation at temperatures even higher than the maximum ambient in order to compensate for extra losses produced by the higher forward voltage drop found in this class of materials. Rectification at temperatures above the 315 C ambient has been observed in gallium arsenide, gallium phosphide, and silicon carbide. Gallium phosphide has been given prime consideration for this particular application because it has a relatively low forward drop of about 2 volts at 500 C, its maximum operating temperature is less limited than is the gallium arsenide which has a ceiling of about 400 C, and when compared with silicon carbide, it has a lower forward drop and better low-temperature performance.

The semiconducting compound gallium phosphide presents all the production problems involved in making device-grade germanium and silicon rectifiers, plus the additional problems that two elements rather than one are involved and the high-temperature reaction pressure at formation of the compound is over 10 atmospheres. The requirements for large crystals with a high degree of perfection to be formed from two very pure elements with exact stoichiometry of the compound call for extreme care throughout the manufacturing processes represented by the flow chart appearing in Fig. 14. Equipment used for growing single crystals from a pool of molten gallium phosphide is shown in Fig. 15. The induction heating coil is at the bottom of the enclosure, and the magnetic suspension drive is visible at the top of the window. The desired degree of perfection has not yet been achieved in all the steps for production of gallium phosphide rectifiers. However, the form of a finished unit is shown by the phantom view in Fig. 16.

Capacitor

High-temperature capacitors need a dielectric material structurally able to withstand wide temperature changes while maintaining usable dielectric characteristics. Vacuum deposited quartz on thin aluminum foil is a good combination which also has the advantage of producing a low bulk factor in the finished capacitor. A 12-microfarad unit may occupy a volume as small as 2 cubic inches.

The vacuum chamber used to coat foil for production of this type of capacitor is shown in Fig. 17. Inside the cylindrical shield at the bottom is a tantalum crucible from which quartz is continuously evaporated when heated by electron bombardment. To replace quartz evaporated from the crucible, a feed mechanism at the top of the chamber introduces more through a tube, as it is needed. Manipulation of the foil to be double coated is performed by a mechanism in the central rectangular framework. A detail view of it is given in Fig. 18. The mechanical driving of the left-hand coupler is done by a shaft extending through a vacuum seal at the rear of the chamber.

Capacitors are produced by concentric winding of two double quartz coated strips with a slight offset to provide a portion at each end for attaching leads. The motorized winder used for this operation appears in Fig. 19. Two forms of

Fig. 21. Testing high-temperature capacitors

encapsulation for the finished capacitors are shown together with a scale to gage their size in Fig. 20. The vacuum-tight metal enclosures have high-alumina insulated terminals for connection of external leads. Checking capacitance of finished units over a wide temperature range is done with the bridge-type instrument appearing at the left of Fig. 21 while connected to the bank of four thermally insulated test chambers. Over the −50 to +450 C range, the capacitance temperature coefficient is about 0.01% per degree centigrade.

Conclusion

The evaluation of these materials and components indicates a successful transformer-rectifier power supply can be built with the circuit configuration described.

Trends in Ground Bed Design for Cathodic Protection of Underground Structures

A. P. LANDRY
ASSOCIATE MEMBER AIEE

I. N. HOWELL, JR.
ASSOCIATE MEMBER AIEE

CATHODIC protection of underground structures for the mitigation of corrosion to those structures has long been an established means for controlling corrosion. It is interesting to note that this paper deals primarily with a relatively new technique in the field of corrosion, proposed originally by the man given credit in "Who's Who" for the development of cathodic protection. This man, Mr. R. J. Kuhn, a consulting engineer in New Orleans, La., has made several contributions in the field of corrosion control. The original idea of cathodic protection is one of impressing a direct current in such a way as to counteract the tendency of the galvanic or man-made electric currents which tend to leave the underground structures, thereby counteracting the loss of metal through the electrochemical process known as corrosion or electrolysis.

Cathodic protection of underground structures using galvanic anodes composed of zinc or magnesium has many applications in industry today. When protection requirements of the structure exceed that available from galvanic anodes, a method using an impressed current must be employed. The anode material can then be any expendable metal desired. The conventional distributed anode type of ground bed, in which manufactured anodes are placed in a suitable horizontal array near the surface of the earth, is in wide use today. Anodes commonly used in the conventional ground bed are graphite and high-silicon cast iron.

Southern Bell Telephone and Telegraph Company installed the first ground bed for protection of underground cable sheaths in New Orleans in 1931. It was a conventional ground bed using graphite anodes. Seven more conventional ground beds were installed in the period from 1931 to 1951.

Since Southern Bell has invested more than five million dollars in underground cables in the New Orleans area, and 92% of the cables have lead sheaths, the problem of cathodic protection is an important one. Fig. 1 shows the number of sections of underground cable which had to be replaced in the period from 1951 to 1955. These sections failed because of sheath breaks due to corrosion and had to be replaced with new cable. The cost of the replacement cables is also shown in Fig. 1. Since the total cost of replacing corroded sections exceeded 216,000 dollars for this period, efforts were made to speed up the cathodic protection program as much as possible. Nine more conventional ground beds were installed in this period.

Experience with Conventional Ground Beds

Efforts to speed up the cathodic protection program were hampered by certain characteristics of the conventional ground bed. Considerable engineering and survey time was consumed in finding locations suitable for construction of the ground bed and the obtaining of the necessary right-of-way.

Close co-ordination was essential with other utilities having underground structures in the area. Extensive tests with New Orleans Public Service, Inc., and Sewerage & Water Board of New Orleans were required with each new installation. These joint tests were used to determine the extent of influence from the ground bed, and the amount of drainage required to offset this influence. Costs involved in making the joint tests ran high; as high as 3,900 dollars in total cost on a particular ground bed installation.

Another factor of importance was the fact that so much of the rectifier output was required for drainage on structures other than the one under primary consideration. As much as 36% of the rectifier current output was required to offset undesirable influences to other structures. This meant higher operating costs and shorter life ground beds than otherwise required.

It had been found through experience that if a minimum horizontal separation of 150 feet was maintained between the ground bed and other structures, unde-

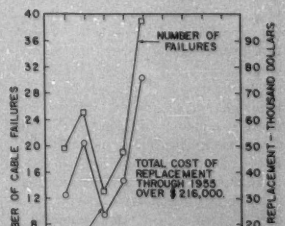


Fig. 1. Cable failures and cost of replacement since 1951 for New Orleans, La.

Paper 59-820, recommended by the AIEE Chemical Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 19, 1959; made available for printing May 22, 1959.

A. P. LANDRY and I. N. HOWELL, JR., are with Southern Bell Telephone and Telegraph Company, New Orleans, La., and Atlanta, Ga., respectively.

The authors wish to thank all who co-operated in the preparation of this paper.

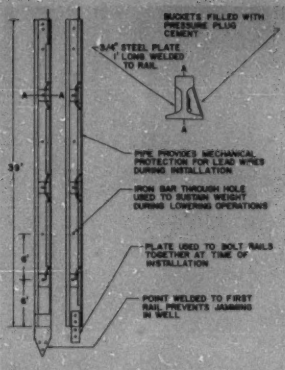

BUCKETS FILLED WITH PRESSURE PLUG CEMENT
3/4" STEEL PLATE 1' LONG WELDED TO RAIL
PIPE PROVIDES MECHANICAL PROTECTION FOR LEAD WIRES DURING INSTALLATION
IRON BAR THROUGH HOLE USED TO SUSTAIN WEIGHT DURING LOWERING OPERATIONS
PLATE USED TO BOLT RAILS TOGETHER AT TIME OF INSTALLATION
POINT WELDED TO FIRST RAIL PREVENTS JAMMING IN WELL
39'
6'
6'

Fig. 2. Rail anode fabrication details

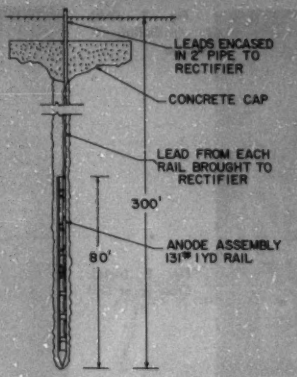


Fig. 3. Deep ground bed installation

sirable influences could be reduced to negligible values with a drainage current not to exceed 10 amperes. It is, of course, obvious that this separation, while being sufficient in New Orleans, may not be enough in other areas. These criteria would, of necessity, have to be determined for any specific application by tests. Serious consequences could otherwise result. These separation criteria were substantiated by joint tests made by Southern Bell and New Orleans Public Service, Inc., in 1956. The results of these tests are shown in the Appendix.

During 1955 and early 1956, a total of 13 additional shallow ground bed sites were located. This seemed to have exhausted the supply of available locations.

Deep Vertical Ground Bed

The idea for the deep vertical type of ground bed was obtained from Mr. R. J. Kuhn, Consulting Engineer, and Mr. S. E. Trouard,[1] New Orleans Public Service, Inc. The deep ground bed had been used successfully for cathodic protection of steel gas mains in areas where conventional ground beds proved impracticable. The ground bed as used by New Orleans Public Service, Inc., was essentially composed of used rail in a well 8 inches in diameter by 250 feet deep.

The basic design as used by New Orleans Public Service, Inc., was adapted for application to Southern Bell requirements. The size of the anode material was increased to compensate for larger current drains (about 10 amperes) required. Other changes were made to facilitate the fabrication and handling of the anode by a local contractor.

Fig. 2 shows general fabrication details of the rail anode assembly as used by Southern Bell. Two 39-foot used rails weighing 130 pounds per yard are employed. One anode lead from each rail is brought to the ground surface. Each lead is attached to the rail in three places. Each point of attachment is protected by a welded bucket filled with pressure plug cement, and a tar sealing compound is poured after heating. Each lead is threaded through metallic pipe welded the length of the rail for mechanical protection during installation. The lead from the first rail is threaded through the pipe on the second or top rail at the time of installation.

A cone-shaped point is welded to the first rail to avoid possible jamming in the well during the lowering operation. The two rails are bolted together with steel joint plates at the time of installation.

Fig. 4. Rail anode showing cone point

Fig. 5. First rail connection

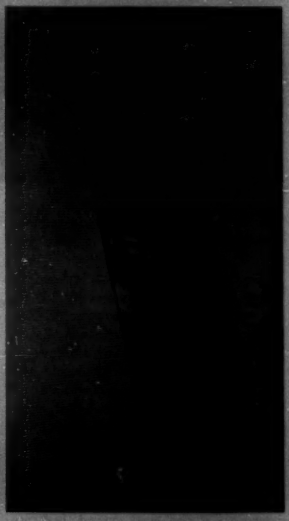


Fig. 6. Second rail connection

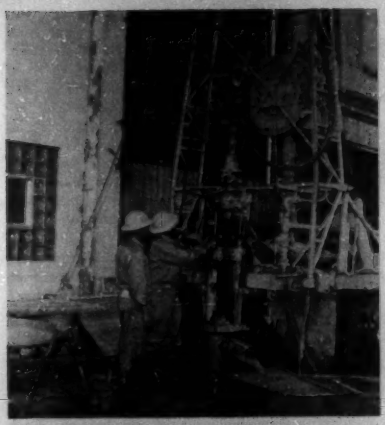


Fig. 7 (left). Drill rig

Fig. 8 (right). Close-up of drill rig

Fig. 9 (left). End of lowered first rail

Fig. 10 (right). Joining rails

This forms a completed anode assembly approximately 80 feet long, weighing about 3,500 pounds.

The completed assembly is lowered into a well 10 inches in diameter by 300 feet deep drilled by a local contractor. Fig. 3 shows a typical ground bed installation. Since the well has no casing, the lowering operation takes place immediately after drilling is completed. The hole will usually remain open somewhat in excess of 6 hours which permits adequate lowering time. The assembly is lowered with galvanized steel strand which is cut off and dropped into the well after completion. Tests were made in 1956 to determine if the steel strand had any undesirable influence; none was detected.

No special backfill is used in addition to the regular amount of drilling mud used by the contractor in his drilling operations. The hole closes in of its own accord in a day or so.

Figs. 4–11 show equipment and installations. The estimated useful life of this ground bed is calculated on the basis of failure after two thirds of the metal is consumed. For current densities used (8 to 10 amperes per ground bed), a useful life of 10 to 13 years is expected.

Total cost of the installed anode assembly is approximately 1,200 dollars. The cost of a shallow ground bed of equal life would cost approximately 1,000 dollars, plus an undetermined amount spent on co-ordinated tests.

A cost breakdown showing the various costs involved in fabrication and installation of the deep vertical ground bed is shown in the Appendix.

Results Using Vertical Ground Bed

Fig. 12 shows the number of cable failures since 1951. The curve shows a marked decline in the period after 1955. The number of failures has decreased from 39 in 1955 to a total of 3 for 1958. The use of the deep ground bed has enabled the cathodic protection program to proceed with minimum delay and achieve these results.

Ground bed resistances have been quite satisfactory. There has been no

Fig. 11 (left). Rectifier installation

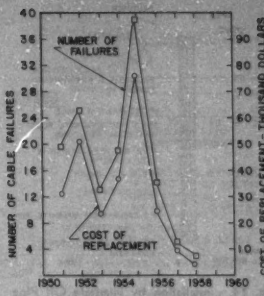


Fig. 12 (right). Cable failures and cost of replacement

evidence of gas blocking in approximately 3 years of operation. Table I shows some typical values of over-all circuit resistance for conventional and deep ground beds.

Ground bed resistances in New Orleans have been extremely satisfactory. It would be appropriate, when locating beds of this type in other areas, to make soil resistivity tests by vibro-ground, earth resistivity megger, or other satisfactory means to the expected drilling depth so that depth selection might be made to facilitate an economical and satisfactory installation. No tests of this type were made, however, in the New Orleans installations since it was well known from existing data what the results should be.

Table I. Typical Circuit Resistances for Conventional and Deep Ground Beds

Ground Bed Designation	Type	Over-all Circuit Resistance (Ohms)
Richard & Eighth	Conventional	1.28
Franklin & Perry	Conventional	0.62
5th & Ave. "K"	Conventional	1.33
Aviators & Bancraft	Conventional	0.89
Clara & Jena	Conventional	1.07
Ave. "K" & 8th	Conventional	1.73
St. Charles C.O.	Conventional	0.80
Rayne & Van	Conventional	1.55
5th & Belanger	Conventional	1.15
Toulouse & Bourbon	Conventional	1.17
Maple & Oak	Deep	0.46
Coliseum & Joseph	Deep	0.39
Dorgenois & Columbus	Deep	0.35
Monroe & Solon	Deep	0.45
Coliseum & Louisiana	Deep	0.38
Burgundy & Floor	Deep	0.39
Painters & Benefit	Deep	0.31
Romap & Spain	Deep	0.36
Marais & Montegut	Deep	0.33
Erato & Roman	Deep	0.62

Deep Ground Bed Using High-Silicon Cast Iron

A new anode assembly has been developed and is being installed by Southern Bell using high-silicon cast iron. The new assembly is designed with a useful life of indefinite length (calculated in excess of 100 years). The cost of the completed assembly is 45% that of the rail and total weight is 500 pounds or 15% the weight of the complete rail assembly.

The fabrication details are shown in Fig. 13. Four 3-inch-diameter 60-inch-long high-silicon cast iron anodes are strapped to a length of channel iron which has been painted with a coal tar enamel. The channel iron is intended to hold the assembly in a vertical array during lowering and until the hole settles in around the anodes. The assembly is lowered into a 10-inch-diameter 200-foot-deep well, with a 1/2-inch manila rope. This avoids placing undue strain on lead connections and also insures electrical remoteness of the ground bed.

The anode assembly using high-silicon cast iron appears to offer a ground bed of equal electrical performance at lower initial cost, lower maintenance costs, and much longer useful life. From the data available at the present time, it appears that the high-silicon cast-iron assembly has advantages which make it preferable to the rail assembly.

Other Deep Ground Bed Designs

Interstate Oil Pipe Line Company in Shreveport, La., has developed and has used a deep ground bed of graphite rods with backfill at depths up to 250 feet. General details are shown in Fig. 14. Thirty-three, 3-inch by 60-inch graphite rods are assembled to form a continuous anode 175 feet in length. The anode is centered in an $8^5/_8$-inch casing filled with coke-breeze backfill. The anode leads are brought to the surface; one from the top of the graphite anode assembly and one from the casing. Expected life is over 20 years at currents of approximately 30 amperes. Ground bed resistances of less than 1 ohm are reported.

Fig. 15 shows how Interstate plans to use a ground bed using high-silicon cast-iron anodes. A coke-breeze backfill will be used.

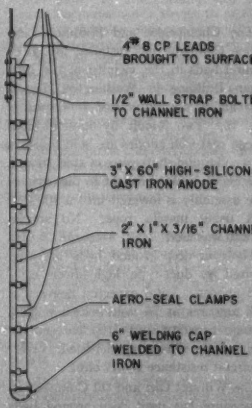

4" # 8 CP LEADS BROUGHT TO SURFACE
1/2" WALL STRAP BOLTED TO CHANNEL IRON
3" X 60" HIGH-SILICON CAST IRON ANODE
2" X 1" X 3/16" CHANNEL IRON
AERO-SEAL CLAMPS
6" WELDING CAP WELDED TO CHANNEL IRON

Fig. 13. High-silicon cast iron anode

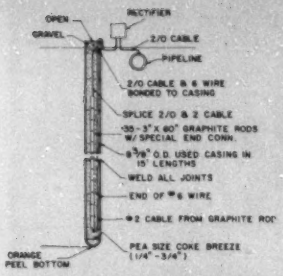

OPEN
GRAVEL
RECTIFIER
2/0 CABLE
PIPELINE
2/0 CABLE & #6 WIRE BONDED TO CASING
SPLICE 2/0 & 2 CABLE
35 - 3" X 60" GRAPHITE RODS W/ SPECIAL END CONN.
6 5/8" O.D. USED CASING IN 10' LENGTHS
WELD ALL JOINTS
END OF #6 WIRE
#2 CABLE FROM GRAPHITE ROD
PEA SIZE COKE BREEZE (1/4" - 3/4")
ORANGE PEEL BOTTOM

Fig. 14. Installation by Interstate Oil Pipe Line Company using graphite anodes

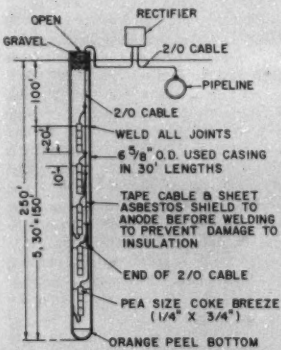


Fig. 15. Installation by Interstate Oil Pipe Line Company using high-silicon cast iron anodes

The Chesapeake and Potomac Telephone Company has installed four deep ground beds in the vicinity of the Pentagon Building in Washington, D. C. The anode assembly used is shown in Fig. 16. Two 3-inch by 60-inch high-silicon cast-iron anodes are held together by a collar and short piece of steel strand. One anode lead is brought to the surface. The assembly is lowered into a 100-foot well using manila rope. No special backfill was used.

The four deep ground beds are augmented by duct-type high-silicon cast-iron anodes. Total expected delivery is 32.7 amperes at 25 volts which is 83% voltage capacity and 71% current capacity of the silicon rectifier used. Overall circuit resistance is 0.77 ohm.

The Willmut Gas and Oil Company in Mississippi has used deep ground beds backfilled at depths of 350 feet.[3] The anode assembly is made up of high-silicon cast-iron anodes attached to a 1-inch metallic pipe used to lower the anodes into the well. A coke-breeze backfill is pumped through the pipe until it flows from the top of the hole. The pipe is then removed by means of a loose coupling at the anode assembly to achieve electrical remoteness of the ground bed. Two main header wires are brought to the surface. Anode to earth resistances of less than 1 ohm have been achieved with no evidence of gas blocking after 1 year of operation.

The method developed by Steele and Associates, Inc., of Atlanta, Ga., makes use of high-silicon cast-iron anodes in a 6-inch-diameter by 200-foot-deep well. The anodes are strapped to a 3/4-inch plastic pipe as shown in Fig. 17. The plastic pipe is used to lower the anodes and to pump backfill of standard drilling mud, bentonite clay, aqua gel, or equivalent, into the well. The plastic pipe, which assures electrical remoteness of the ground bed, is left in the well in the event it becomes necessary to break gas blocking at some future date. This remoteness which is achieved is a relative quantity and refers to other installation techniques which may involve the use of metallic stranded cable or metallic pipe for lowering the assembly into place. Steele and Associates feel that this plastic pipe, which is left in the ground, may be used to flush the anode installation for purposes of breaking any gas blocking should it occur at some future date. Some 50 installations of similar nature have now been installed by these engineers.

Conclusions and Trends

Many problems which have been encountered when applying cathodic protection to underground structures have been alleviated by the use of the deep vertical ground bed. Extensive, coordinated tests with other utilities were made unnecessary by mutual agreement since the adverse influence on other structures are undetectable with present instrumentation. The deep ground bed using high-silicon cast iron has a very long useful life, and does away with expensive ground bed replacements. The use of backfill in higher resistance soils has been successful in lowering ground bed resistances to quite acceptable values.

As more experience is gained with the newer designs of deep ground beds, it is expected that improvements and refinements will be made. As improved anode materials and drilling techniques

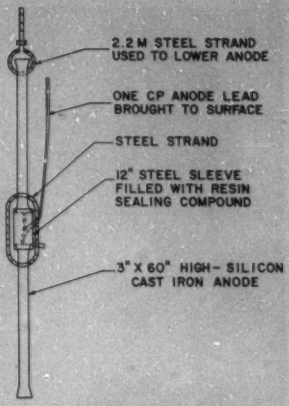

2.2 M STEEL STRAND USED TO LOWER ANODE
ONE CP ANODE LEAD BROUGHT TO SURFACE
STEEL STRAND
12" STEEL SLEEVE FILLED WITH RESIN SEALING COMPOUND
3" X 60" HIGH - SILICON CAST IRON ANODE

Fig. 16. Anode assembly used by the Chesapeake & Potomac Telephone Company

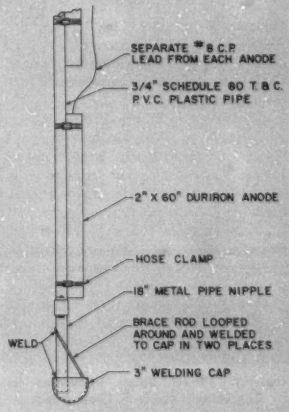

SEPARATE #8 C.P. LEAD FROM EACH ANODE
3/4" SCHEDULE 80 T. & C. P.V.C. PLASTIC PIPE
2" X 60" DURIRON ANODE
HOSE CLAMP
18" METAL PIPE NIPPLE
BRACE ROD LOOPED AROUND AND WELDED TO CAP IN TWO PLACES
WELD
3" WELDING CAP

Fig. 17. Steele & Associates, Inc., anode assembly

are developed, deep ground beds should become attractive for application of cathodic protection where it is not practical at the present time. The deep ground bed has answered a real need in New Orleans and will continue to serve well in the years to come.

Appendix

The following table lists the cost of major items involved in the fabrication and installation of the deep ground bed using rails. Figures are approximate and reflect the

average of figures which have been encountered in the past.

Typical Costs of Rail Ground Bed

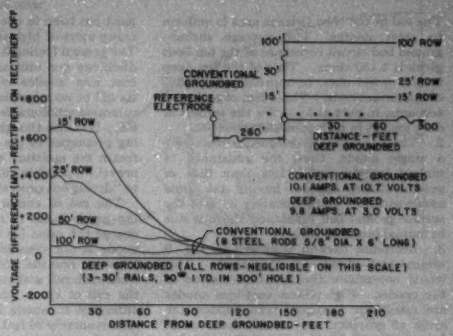
3,300 pounds used steel rail	$ 330.00
700 feet no. 6-type *RR* wire	210.00
Fabrication charges	170.00
350 feet 10-*M* galvanized steel cable	30.00
Pipe, miscellaneous steel, etc.	70.00
Labor and equipment for lowering anode into well	90.00*
Anode well (10 inches diameter by 300 feet deep)	300.00
Total	$1,200.00

* This item is variable over wide ranges due to dependence on completion of drilling operation by well contractor.

Fig. 18 shows a plot of the results obtained in tests performed jointly by Southern Bell Telephone and Telegraph and New Orleans Public Service, Inc. As indicated in Fig. 18 two ground beds were installed in a vacant lot; one a deep ground bed using a rail anode assembly, the other a test ground bed composed of steel rods 5/8 inch diameter.

Voltage gradient measurement were made using two copper sulfate half-cells. Readings were taken with one half-cell held stationary at a remote location while the other half-cell was moved to various locations throughout the area. Readings were recorded with the rectifier on and with rectifier off for both the conventional ground bed and the deep ground bed. The difference between the "rectifier on" readings and "rectifier off" readings is plotted as a function of distance in Fig. 18. The data recorded for the deep ground bed did show some slight variations, but they were negligible when plotted on the scale used.

Fig. 18. Plot of results using two ground beds



The change in earth potential due to the conventional ground bed is appreciable up to distances of 150 feet, while the changes due to the deep ground bed are negligible at all distances.

References

1. Cathodic Protection of the Coated Steel Gas Main Distribution System in New Orleans, S. E. Trouard. *Corrosion*, Houston, Tex., vol. 13, no. 3, Mar. 1957, pp. 31-31.

2. Impressed Current Anodes Installed and Backfilled at Depth of 350 Feet, J. F. Tatum. *Corrosion*, Houston, Tex., vol. 14, no. 4, Apr. 1958, pp. 98-100.

3. Deep Ground Beds for Line Protection, R. L. Bullock. *Pipe Line Industry*, Houston, Tex., vol. 9, no. 4, Oct. 1958, pp. 40-44.

Discussion

G. P. Miles (Interstate Oil Pipe Line Company, Shreveport, La.): In reading and commenting on this paper, I found many points of agreement and many new ideas which may be useful in the design of future deep ground bed installations. There are several other points which need clarification and still others with which I am not in agreement.

First, let us consider some of the advantages listed for deep ground beds. The authors listed the elimination of right-of-way problems and extensive co-ordinated tests with other utilities as major advantages. I am in full agreement with the first and in partial agreement with the second.

Though the elimination of right-of-way problems is certainly a major advantage for utility companies in congested areas, an advantage of even greater importance is the adaptability of these ground beds to high resistance soil. The use of deep ground beds has afforded an economical method of applying cathodic protection to bare pipe lines which were heretofore considered impractical to protect due to their location in ultrahigh-resistance surface soil. Their adaptability to this type of soil makes it possible to install low-resistance ground beds at any of the scattered locations of electric power along pipe lines. Interstate presently has 13 cathodic protection rectifier-deep ground bed units in service. Many of these installations are in soil of 100,000 ohms per centimeter and all are operating at below 1 ohm total circuit resistance with the exception of one which is 3 ohms.

Another extremely important advantage of deep ground beds is the extensive protective current spread obtained. Protection from deep ground beds in many cases is equal to three times the pipeline surface area which could be protected from a normal ground bed with equal current output.

In regard to the elimination of extensive interference surveys, interference is minimized, but the need for such surveys is not eliminated. Remember that all ground beds act differently and the elimination of interference surveys based on the results of one or a 100 tests can mean serious and costly damage. The reported test of interference using two copper sulfate half-cells leaves much to be desired. Interference tests conducted in areas where potentially interfering metallic conductors are not present seem baseless. Possibly, these tests were performed with interfering elements present; however, the magnitude of potential damage from improperly bonded interfering metallic masses is such that this point cannot be taken lightly.

Two excerpts from the Interstate papers referenced by the authors need comment. First, it was stated that the life of Interstate's deep ground beds is over 20 years. This is correct though the 20 years, which is a conservative estimate, is based on a current output of from 40 to 60 amperes rather than the 30 amperes stated.

Further, it was correctly stated that Interstate plans to use a deep ground bed using high-silicon cast iron anodes. This ground bed has been installed. However, it was of an experimental nature. Due to the erratic results reported for this material by others, we plan to proceed cautiously, watching results, evaluating data, and proving their worth before proceeding with additional ground beds of this material.

While on the subject of high-silicon cast-iron anodes, the reference to a ground bed life "calculated in excess of 100 years" is certainly something to think about. Using a manufacturer's literature, I once calculated a normal ground bed life in excess of 50 years for an area where galvanic anodes were being expended in 3 to 4 years. These calculations were filed away without being quoted, with the notation "FANTASTIC." Aren't we buying a pig in a poke or trying to make good on the lifetime guarantee of a fountain pen, when we let our enthusiam run away with us?

This paper by Mr. Landry and Mr. Howell has performed a service to pipeline and utility-corrosion engineers by combining many deep ground bed designs in current use into one composite paper. It is my opinion that the widespread use of the deep vertical-type cathodic protection ground bed has been one of the most significant developments in the field of cathodic protection during the last 5 or more years. The further exploration and development of this useful tool of corrosion control offers a challenge to all whose job is the protection of buried metallic structures.

Lyle R. Sheppard (Shell Pipeline Corporation, Houston, Tex.): The paper by Landry and Howell on deep ground beds is a very good summary of the recent work. However, I believe several points of ground bed design have not been pointed out, and thus without their consideration, the anticipated favorable results may not always be attained.

1. An anode at the surface of the earth has approximately twice the electrical resistance to earth as the same anode, in the same resistivity soil, at great depths in the earth.

The soil in the New Orleans area is uniform to great depths. The average surface-ground bed circuit resistance of the ten beds given is 1.159 ohms. That of the ten deep beds is 0.404 ohm, or approximately 35% of the surface beds. This effect of geometry accounts, in part at least, for the success of deep anodes in high resistance soil.

2. The deep anode as designed is effectively a single anode, thus, the resistance of mutual interference is less than that of several shallow anodes having the same anode surface. Approximately 30% of the surface ground bed resistance above is due to mutual interference.

3. Bentonite backfill is shown in Fig. 9 around high-silicon cast-iron anodes. This would accentuate the surface corrosion of the anodes. High-silicon cast-iron anodes are essentially "oxygen"-type anodes, the same as graphite, and should have coke-breeze backfill to take the anode corrosion away from the anode surface.

4. The resistances of the individual leads to various parts of the deep anode should be equalized by insertion of additional resistances, in order that the anode will be uniformly consumed without "necking."

A. P. Landry and **I. N. Howell, Jr.**: The authors are grateful for Mr. Sheppard's and Mr. Miles' comments and additional points of ground bed design contained in their discussions. In reference to Mr. Sheppard's comments, the phenomenon of the anode to earth resistance varying with the anode depth has been discussed in the literature.[1] Bentonite backfill referred to by Mr. Sheppard has been in use for some time now by many users of high-silicon cast iron anodes. The general feeling is that as long as current discharge per unit area of the high-silicon cast iron anodes is not excessive, there should be no appreciable increase in surface corrosion attributable to the bentonite backfill. It is also thought that the water holding properties of the bentonite will tend to retain the moisture around the anode and prevent a possible increase in resistance due to "drying out process" around the anode.

The use of compensating resistances in the anode leads to make the current discharge the same for all sections of the rail anode is not considered necessary for the following reasons. Since a lead is brought from each rail and each lead is attached to the rail at three locations, the maximum difference in anode lead length would be approximately 24 feet. The resistance of 24 feet of no. 6 wire is of such small proportions, its effect can be safely neglected. An additional factor to consider is that if necking does occur, the three lead contact points will insure continued consumption of metal with little increase in circuit resistance.

Mr. Miles has made reference in his discussion to the interference tests described in the paper. The results of these tests, which were made in the low-resistance soil in New Orleans and shown on the potential profile plot (Fig. 18), indicated that relatively steep potential gradients are produced by the conventional ground bed. The data taken for the deep ground bed show no appreciable potential gradient. No excessive current flow can be induced on an interfering metallic structure if there is no appreciable potential gradient to cause this current flow. Other tests not discussed in the paper have been conducted for the expressed purpose of determining the influence of a deep ground bed (operating at approximately 10 amperes) on all underground metallic structures in the vicinity. These tests, without exception, indicated no appreciable adverse effects on any underground structure.

The deep ground bed using high-silicon cast iron anodes has been modified slightly, eliminating all local fabrication. Each anode is lowered into the well using its individual lead and is suspended at the desired depth from this lead. This greatly facilitates and expedites the placing of the ground bed by the contractor. It was not the authors' intent to convey the impression that "indefinite life" was the main consideration in the use of high-silicon cast iron for the anode material. Other equally important considerations are reduced cost, reduced weight and size, and elimination of local fabrication intervals. At the same time, published laboratory and field measurements indicate that a service life can be expected which far surpasses the expected life of the rail anode. The actual extent to which anode life will be extended can be determined only by the results of data obtained over a long period of time.

REFERENCE

1. DEEP GROUND BEDS FOR CATHODIC PROTECTION, S. E. Troutard, E. A. Wagner, Jr. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, Nov. 1959, pp 278–85.

Temperature Prediction in Thermal-Lag Equipment

P. B. RICHARDS
NONMEMBER AIEE

IN MISSILE applications, electric equipment is required to function for a relatively short period of time in a hot ambient atmosphere, but is deprived of external cooling media such as air or oil. The dependability of the equipment depends on its ability to absorb its own heat losses or to transmit these losses by conduction to cooler surroundings. The general purpose of this paper is to present a method for predicting simultaneously the transient temperatures in two heat-generating solids such that both materials store some heat, while some heat is transmitted from one to the other and thence possibly to a third region at a fixed temperature. In particular, mean transient temperatures are predicted in the field winding and laminations of an a-c generator which is bolted to a support, but otherwise insulated from the ambient atmosphere. Application of this method to temperature prediction in solid-state devices is natural.

Paper 59-913, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing May 4, 1959.

P. B. RICHARDS is with the General Electric Company, Philadelphia, Pa. This paper is the outgrowth of work performed while the author was with the Research and Development Department of Jack and Heintz, Inc.

Statement of Problem

The significant heat losses in the rotor (Fig. 1) are the winding I^2R loss in the coil and the pole face loss in the iron. It will be assumed that no heat is transmitted across the air gap, hence the aforementioned losses must be stored either in the rotor coil and iron or transmitted along the shaft to the mounting pad (support). The machine is completely insulated from the hot atmosphere by necessity, so the mounting pad is the only available heat sink.

The energy balance can be expressed by the following equations

$$Q_c = Q'_c - Q_{ci} \tag{1}$$

$$Q_i = Q'_i + Q_{ci} - Q_{ip} \tag{2}$$

where

Q_c = heat absorbed by coil
Q'_c = heat generated by coil (I^2R loss)
Q_{ci} = heat conducted from coil to lamination
Q_i = heat absorbed by lamination
Q'_i = heat generated in lamination (pole face loss)
Q_{ip} = heat conducted from lamination to pad

These quantities of heat are expressible in terms of the time t, mean temperature $T(t)$, mass M, specific heat C, and thermal resistance R as follows:

$$Q_c = M_c C_c \frac{dT_c}{dt} \tag{3}$$

$$Q'_c = A + B\,T_c,\ A,\ B \text{ constants} \tag{4}$$

$$Q_{ci} = \frac{T_c - T_l}{R_{ci}} \tag{5}$$

$$Q_l = M_l C_l \frac{dT_l}{dt} \tag{6}$$

$$Q'_l = Q_{PF} = \text{constant} \tag{7}$$

$$Q_{lp} = \frac{T_l - T_p}{R_{lp}} \tag{8}$$

where the subscripts c, l, p denote coil, laminations, and pad respectively, R_{ci} is the thermal resistance between the coil and the laminations, R_{lp} the thermal resistance between the laminations and pad, and Q_{PF} the pole face losses.

Substituting these expressions in the energy balance formulas, two simultaneous differential equations are obtained for $T_c(t)$ and $T_l(t)$:

$$M_c C_c \frac{dT_c}{dt} = A + BT_c - \frac{(T_c - T_l)}{R_{ci}} \tag{9}$$

$$M_l C_l \frac{dT_l}{dt} = Q_{PF} + \frac{T_c - T_l}{R_{ci}} + \frac{T_p - T_l}{R_{lp}} \tag{10}$$

subject to the initial conditions

$$T_c(0) = T_0$$

$$T_l(0) = T_0, \qquad T_0 = \text{constant}$$

Solution

Equations 9 and 10 are readily solved by the Laplace transform method, assuming the material thermal properties are constant. The procedure will be explained in sufficient detail to permit numerical calculation of solutions.

The equations are first simplified by the notation

$$T_c(t) = y(t) \qquad T_l(t) = x(t)$$

$$a_1 = M_c C_c \qquad a_2 = \frac{1}{R_{ci}}$$

$$a_3 = B - \frac{1}{R_{ci}} \qquad a_4 = A \tag{11}$$

$$b_1 = M_l C_l \qquad b_2 = -\left(\frac{1}{R_{ci}} + \frac{1}{R_{lp}}\right)$$

$$b_3 = \frac{1}{R_{ci}} \qquad b_4 = Q_{PF} + \frac{T_p}{R_{lp}}$$

to

$$a_1 y'(t) - a_2 y(t) - a_3 x(t) = a_4$$

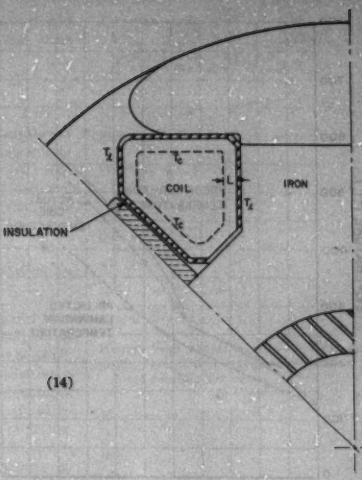
$$b_1 x'(t) - b_2 y(t) - b_3 x(t) = b_4$$

The Laplace transforms of x and y are

$$L\{x\} = \frac{A_1 p^2 + B_1 p + C_1}{A_2 p^3 + B_2 p + C_2} = \frac{N(p)}{D(p)} \tag{12}$$

$$L\{y\} = \frac{A'_1 p^2 + B'_1 p + C'_1}{A_2 p^3 + B_2 p + C_2} = \frac{N^*(p)}{D(p)} \tag{13}$$

Fig. 1. Cross section of rotor



where

$$\begin{aligned}
&A_1 = a_1 b_1 T_0 \\
&B_1 = a_1 b_4 + T_0(a_1 b_2 - a_2 b_1) \\
&C_1 = a_4 b_2 - a_2 b_4 \\
&A'_1 = a_1 b_1 T_0 = A_1 \\
&B'_1 = a_4 b_1 + T_0(a_2 b_1 - a_1 b_1) \\
&C'_1 = a_2 b_4 - a_4 b_2 \\
&A_2 = a_1 b_1 \\
&B_2 = -(a_1 b_2 + a_2 b_1) \\
&C_2 = a_2 b_2 - a_3 b_3
\end{aligned} \tag{14}$$

Using Pipes' transform pair #52,[1] the inverse transforms of equations 12 and 13, the desired solutions $T_c(t)$ and $T_l(t)$, are given by the Heaviside Expansion Formula

$$x(t) = \frac{N(0)}{D(0)} + \sum_{r=1}^{2} \frac{N(p_r)e^{p_r t}}{p_r D'(p_r)} = T_l(t) \tag{15}$$

$$y(t) = \frac{N^*(0)}{D(0)} + \sum_{r=1}^{2} \frac{N^*(p_r)e^{p_r t}}{p_r D'(p_r)} = T_c(t) \tag{16}$$

provided p_r, $r = 1, 2$, are distinct nonzero roots of $D(p) = 0$, and

$$D'(p_r) = \frac{dD}{dp}\bigg|_{p = p_r} \neq 0$$

These temperatures will stabilize with time; i.e., become asymptotic to some finite value, provided the roots p_r of $D(p) = 0$ are real and negative. Requirements for this condition will now be obtained.

Temperature Stabilization Criterion

For

$$D(p_r) = A_2 p_r^3 + B_2 p_r + C_2 = 0$$

the roots are given by

$$p_r = \frac{-B_2 \pm \sqrt{B_2^2 - 4A_2 C_2}}{2A_2} \tag{17}$$

From equations 14 and 11, assuming $M_c \neq 0$, $M_l \neq 0$, $0 < R_{ci} < \infty$, $0 < R_{lp}$, it is apparent that both roots will be real and distinct, for

$$A_2 > 0,$$

$$\begin{aligned}
B_2^2 - 4A_2 C_2 &= (a_1 b_2 + a_2 b_1)^2 - 4a_1 b_1 (a_2 b_2 - a_2 b_4) \\
&= (a_1 b_2 - a_2 b_1)^2 + 4a_1 a_2 b_1 b_3 > 0
\end{aligned}$$

Since there can be no complex roots, any nonstabilization will be of the nature of increasing temperatures.

It is shown in the Appendix that a necessary and sufficient condition for the stabilization of temperature is

$$B < \frac{1}{R_{ci} + R_{lp}} \tag{18}$$

that is, if, and only if, the rate, with respect to temperature, at which heat is generated in the coil is less than the thermal conductance (reciprocal of thermal resistance) from the coil to the pad, then the temperatures in the machine will stabilize with time.

This rate will be the product

$B = I^2 b$

I = current (constant)

b = rate at which electrical resistance R changes with temperature ($R = a + bT_c$)

The temperatures will stabilize at the finite values

$$T_l(\infty) = \frac{N(0)}{D(0)} = \frac{C_1}{C_2}$$

$$T_c(\infty) = \frac{N^*(0)}{D(0)} = \frac{C'_1}{C_2}$$

Calculation of Solutions

Calculation of equations 15 and 16 is straightforward and easy to program for a

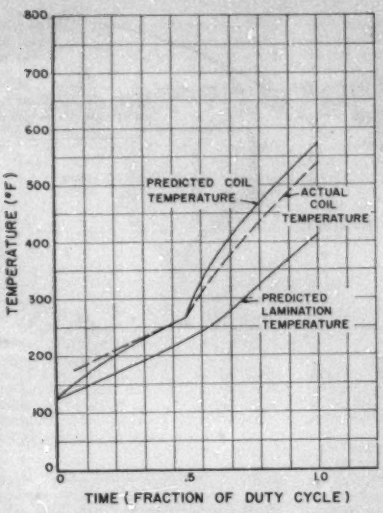


Fig. 2 (left). Temperatures in thermal-lag generator. Initial temperature 125 F. Pad temperature: 125 F, first half of duty cycle; 150 F, final half of duty cycle

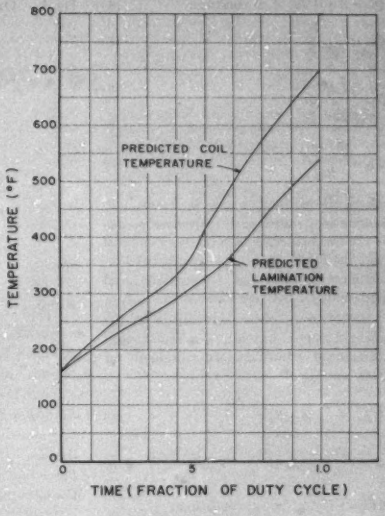


Fig. 3 (right). Temperatures in thermal-lag generator. Initial temperature: 160 F. Pad temperature: 250 F, first half of duty cycle; 325 F, final half of duty cycle

digital computer once the constants in equations 14 and 11 are determined. Of these, only R_{ci} is likely to cause trouble in a rotating machine.

By definition (equation 5), R_{ci} is the thermal resistance between the isotherm which denotes the average temperature T_c in the coil and the isotherm which represents the average temperature T_l in the laminations. For all practical purposes, however, the resistance within the lamination is negligible,[3] so the isotherm T_l coincides with the boundary separating the coil from the laminations (Fig. 1). The coil is surrounded by electrical insulation whose thermal resistance is easy to compute. The remaining resistance is that in the coil between the isotherm T_c and the surrounding insulation. Denote this resistance by R_c. Then

$$R_c = \frac{L}{KA}$$

where L is the distance from this isotherm T_c to that portion of the outer edge of the coil through which heat is transmitted, K is the thermal conductivity of the coil,

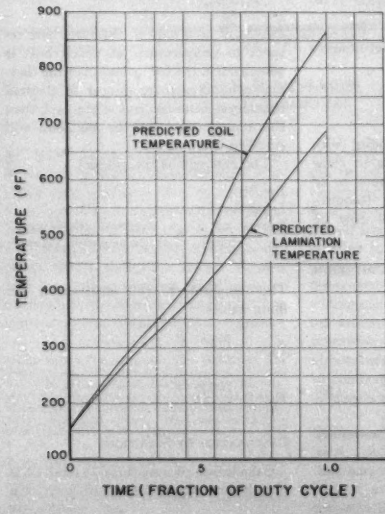


Fig. 4 (left). Temperatures in thermal-lag generator. Initial temperature: 160 F. Pad temperature: 500 F

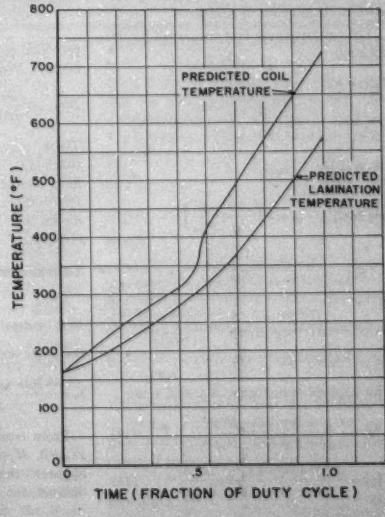


Fig. 5 (right). Temperatures in thermal-lag generator. Initial temperature: 160 F. Generator perfectly insulated from pad

and A is the surface area through which heat is transmitted from the isotherm T_c to the edge of the coil. Equations 3 and 4 suggest that the distance L be defined as follows:

Consider the entire mass M_c of the coil concentrated at the isotherm T_c at a distance L from the outer heat transfer surfaces of the coil such that the total moment $L\,M_c$ is equal to the total moment of the individual masses in the coil with respect to these same surfaces. Since the coil is homogeneous, picture the coil as consisting of n equal masses m_k. Then $M_c = n\,m_k$ and

$$L = \frac{\sum_1^n x_k m_k}{M_c} = \frac{\sum_1^n x_k}{n}$$

where x_k is the shortest distance from any point, mass, in the coil to the heat transfer surface. L can be calculated by placing a small grid over a print of the coil cross section and simply counting the number of nodal points at one unit distance from the heat transfer surface, then the number at two units, and so on.

Using this method of calculating L, solutions of equations 15 and 16 have been computed on the International Business Machines Corporation *650* digital computer for a pad-mounted a-c generator operated under typical missile conditions: half-load for the first half of the duty cycle, followed by full-load for the last half of the duty cycle. The generator is insulated from the pad, drive end, by gasket material. The theoretical solutions are plotted in Fig. 2 as predicted coil and predicted lamination temperature, under the conditions

$T_0 = 125$ F (degrees Fahrenheit), initial temperature
$T_p = 125$ F, first half of duty cycle
$T_p = 150$ F, last half of duty cycle

To verify this analysis, a generator was tested under the same conditions as were used in obtaining the predicted temperatures. The average coil temperature was calculated from the measured field resistance. This temperature is plotted in Fig. 2 as actual coil temperature. Agreement between theory and experiment is within 6%. Satisfied that this theoretical analysis is sufficiently accurate to predict temperatures, a series of solutions were calculated to determine the effect of pad temperature.

Fig. 3 indicates the predicted coil and lamination temperatures for the generator just discussed, but with

$T_0 = 160$ F, initial temperature
$T_p = 250$ F, first half of duty cycle
$T_p = 325$ F, last half of duty cycle

Fig. 4 indicates the predicted coil and lamination temperatures for the same generator, but with

$T_0 = 160$ F, initial temperature
$T_p = 500$ F, entire duty cycle

indicating that with a pad temperature of 500 F, coil temperature is dangerously high for the final 20% of the duty cycle. Fig. 5 indicates the predicted coil and lamination temperatures for the same generator, but with

$T_0 = 160$ F, initial temperature
$R_{1p} = \infty$, generator perfectly insulated from pad

These temperatures are 130 F cooler than when the pad is at 500 F, and only about 30 F hotter than when the pad is at 325 F. One would thus expect that insulation is the solution to high pad temperatures, but this is not practical. Mechanical requirements limit the amount of insulation that can be used between the generator and the pad, while a very large increase in insulation is required to appreciably reduce the coil temperature.

Summary

Two simultaneous differential equations are derived and solved for the transient temperatures in two heat-generating solids such that both materials store some heat, while some heat is transmitted from one to the other and thence possibly to a third region at a fixed temperature. Conditions are determined under which the temperatures will stabilize with time. To demonstrate the theory, temperatures are predicted in the field winding and laminations of an a-c generator which is bolted to a support, but otherwise insulated from the ambient atmosphere. Measured temperatures agree within 6% with predicted temperatures.

The theory can be extended to any number of heat-generating regions, and therefore can be applied to solid-state devices.

Appendix. Proof of Stabilization Criterion

If

$$B < \frac{1}{R_{c1}+R_{1p}}$$

then

$$B < \frac{1}{R_{c1}}$$

and

$$B_2 = -a_1b_2 - a_2b_1$$
$$= M_cC_c\left(\frac{1}{R_{c1}}+\frac{1}{R_{1p}}\right) - M_1C_1\left(B-\frac{1}{R_{c1}}\right) > 0$$

Also

$$C_2 = a_2b_2 - a_2b_2$$
$$= -\frac{[B(R_{c1}+R_{1p})-1]}{R_{c1}R_{1p}} > 0$$

then

$$A_2C_2 > 0$$

and

$$B_2 > \sqrt{B_2^2 - 4A_2C_2}$$

Hence, from equation 17, it is clear that $\rho_r < 0$, $r = 1, 2$ and stabilization follows.

Next consider the case

$$B > \frac{1}{R_{c1}+R_{1p}}$$

then

$$C_2 < 0$$

hence

$$A_2C_2 < 0$$

and

$$|B_2| < \sqrt{B_2^2 - 4A_2C_2}$$

Hence from equation 17 it is clear that one of the roots ρ_r will be positive, and the temperatures will never stabilize.

This will occur, for example, if the machine is perfectly insulated from the pad ($R_{1p} = \infty$), since, in general, B is a positive quantity.

Finally, if

$$B = \frac{1}{R_{c1}+R_{1p}}$$

it is readily apparent that $C_2 = 0$, implying, by equation 17, that one root ρ_r will vanish. This necessitates another form for the solutions than the Heaviside Expansion Formula (equations 15 and 16). Indeed, for $C = 0$, instead of equation 12, write

$$L\{x\} = \frac{A_1\rho_r + B_1\rho}{\rho(A_1\rho+B_2)} + \frac{C_1}{\rho(A_1\rho+B_2)}$$

and by transform pair #9,[1] the inverse transform of x diverges linearly with time. The same will be true for $L\{y\}$. Thus the stabilization criterion equation 18 is established.

References

1. Applied Mathematics for Engineers and Physicists (book), L. A. Pipes. McGraw-Hill Book Company, Inc., New York, N. Y., 1946.

2. Progress in the Thermal Design of Oil-Cooled Rotating Electrical Machines, P. B. Richards. *AIEE Transactions*, pt. II, (*Applications and Industry*), vol. 77, Nov. 1958, pp. 330-34.

Application of Switching Transistors and Saturable Reactors in a High-Performance Servo

FRED B. COX
STUDENT MEMBER AIEE

PAUL R. JOHANNESSEN
ASSOCIATE MEMBER AIEE

THE PURPOSE of this paper is to demonstrate that the performance of instrument servos can be improved by the exclusive use of magnetic devices and transistors to perform the functions of amplification, modulation, and demodulation. Substitution of solid-state devices for electronic amplifiers and mechanical choppers not only increases the efficiency, the reliability, and the durability of the system but also minimizes its size and weight.

Successful use of magnetic amplifiers as servo units has been reported in numerous papers (see, for example, references 1 and 2) but the drift and slow response of conventional 60- and 400-cps (cycles per second) amplifiers prohibit their use in many applications. Furthermore, it is customary to drive the magnetic stage with an electronic preamplifier in order to achieve adequate gain because of the drift problems encountered in cascading two magnetic-amplifier stages. The use of a high-frequency supply voltage has been shown to overcome many of the inherent limitations of a magnetic amplifier.[3] The push-pull amplifier, however, is electrically inefficient because the output circuit dissipates power when the error signal is zero. Power-switching transistors, driven by a magnetic amplifier, have been suggested as a solution to the problem of inefficiency.[4] These ideas, together with many others, however, are as yet unproved as a means of improving over-all servo performance.

The System

A block diagram of the high-performance position servo is shown in Fig. 1, and a schematic diagram of the electric circuits is shown in Fig. 2. The system was designed with no particular application in mind, but every attempt was made to determine an "upper bound" on the performance that might be expected from such a servo. In particular, the system parameters were chosen for maximum bandwidth and velocity constant. Other desirable characteristics, such as negligible drift, short settling time after slew, high efficiency, and insensitivity to small changes in the system carrier frequency follow from the inherent properties of the system components.

The servo amplifier consists of a 2-stage transistor common-emitter voltage amplifier, a magnetic amplifier, and a push-pull switched-transistor power amplifier. The class *A* transistor amplifiers are necessary to achieve the desired loop gain because of the low-voltage amplification of the magnetic and power stages.

The magnetic amplifier is designed to furnish current pulses, the widths of which are proportional to the control signal, to the bases of the power transistors. The magnetic-amplifier supply voltage, furnished by a simple d-c to a-c converter, is a 5-kc square wave, and, because an output pulse occurs on each half-cycle of the supply, the pulse repetition rate is 10 kc. The output polarity reverses on each half-cycle of the 400-cps system carrier. Depending upon the load requirements, the magnetic amplifier must supply from 200- to 500-milliampere peak current to the *H-6* power transistors. The quiescent power dissipation of the magnetic stage is therefore substantially lower than it would be if the magnetic amplifier were used to drive the motor.

Output pulses from the magnetic amplifier control the polarity of the transistor bases with respect to ground, and therefore control the flow of current to the load. Because current flows in the transistors only when an error signal appears in the loop, the efficiency of the power stage is enhanced, and heating in the motor windings is kept to a minimum.

Because it is desirable that the motor control voltage be sinusoidal, a 2-microfarad capacitor is connected across the control windings. The tank circuit, consisting of the capacitor in parallel with the winding inductance and resistance, is tuned for resonance at the 400-cps system carrier frequency. Thus all harmonics of the 10-kc repetition frequency and higher harmonics of the 400-cps carrier are filtered out so that only the information component of voltage is retained.

Minor-loop feedback from an a-c tachometer is used to achieve the desired bandwidth. To obtain a high velocity constant, the tachometer is effectively decoupled from the system for constant motor speeds by a high-pass filter in the feedback loop. To avoid the use of a band-pass filter, the tachometer signal is demodulated prior to filtering by a keyed, phase-sensitive magnetic demodulator. A transistor emitter-follower amplifier is used to isolate the demodulator output.

Fig. 2 shows that position and acceleration errors are summed at the input of the emitter-follower isolation amplifier. It is necessary to tie the two error channels together at this point because of the impedance-matching problems associated with the following stage. The high-pass filter has no effect upon the position error signal, however, because it is a modulated 400-cps signal from a synchro coupled to the motor shaft. Thus, summation of the two voltages is effectively performed at the input of the keyed magnetic modulator.

To sum the a-c position and the d-c acceleration errors, a keyed modulator is used. The modulator samples during a short interval of time (50 microseconds) on each half-cycle of the system carrier. In the interval between samplings, the output voltage is maintained constant by a pulse-stretching circuit. The output, therefore, is a 400-cps square wave whose amplitude is proportional to the difference between the synchro voltage and the filtered, demodulated tachometer voltage.

System Performance

The system described in the preceding section has been built and tested. Results of these tests are summarized briefly as follows:

Velocity constant—approximately 9,500 seconds^{-1}.

Torque constant—0.17 ounce-inch per milliradian, referred to the motor shaft. (This applies to a Kearfott *R901-1A-A* 2-phase motor-rate generator.)

Paper 58-89, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 29, 1957; made available for printing March 31, 1959.

FRED B. COX and PAUL R. JOHANNESSEN are with Massachusetts Institute of Technology, Cambridge, Mass.

The work reported in this paper was supported in part by Project *7668-R*, Servomechanisms Laboratory, Massachusetts Institute of Technology, under contract no. *AF 33(616)-3950* with the Weapons Guidance Laboratory, Wright Air Development Center.

Open-loop gain crossover frequency—230 radians per second.

Settling time for 1-degree step input (referred to the motor shaft)—approximately 20 milliseconds. (Settling time is the time required for oscillation of the response to decrease to less than 5% of the final value.)

The closed-loop sinusoidal frequency response is plotted in Fig. 3. These results were obtained for an input signal corresponding to a 1-degree peak displacement at the motor shaft. Because equipment was not available to obtain modulation frequencies above 50 cps, exact determination of the resonant frequency and the peak magnitude ratio was not possible. The system step response is shown in Fig. 4.

System Components

The units comprising the major components of the servo electrical system are described in the following paragraphs.

Magnetic Amplifier

The magnetic amplifier, shown in detail in Fig. 5, serves as the controller for the push-pull transistor output stage. This amplifier is quite similar to the one described by Fischer[3] and consists of two half-wave phase-reversible push-pull units operating side by side, so that an output

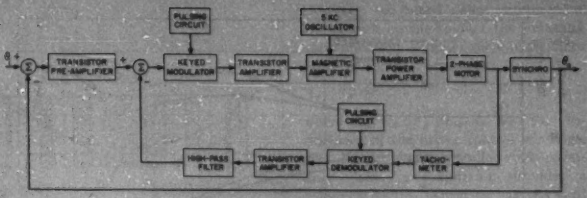


Fig. 1. Schematic diagram of the servo

signal occurs on each half-cycle of the supply voltage.

The supply voltage e_s is 5 kc and square-wave, obtained from a conventional transistor d-c to a-c converter.[1] The high-frequency supply is used in order to minimize the phase lags introduced by the magnetic amplifier. A square-wave supply is recommended rather than a sinusoidal supply because a sinusoidal voltage will allow the output rectifiers to conduct at the beginning and end of each control half-cycle with a resulting loss of gain.

Operation of the unit may be understood with the aid of Fig. 5. From this diagram it is seen that the winding polarities are such that a positive control signal e_c opposes the bias voltage in core IV and aids the bias voltage in core I. Similarly, a positive control voltage opposes the bias in core III and aids the bias in core II. Therefore, when e_{c1} is negative and the associated rectifier is nonconducting, a positive control signal resets the flux of core I to a lower level than that of core IV. The same action takes place between cores II and III when e_{c1} is negative.

When e_{c1} is positive, the flux of cores I and IV is driven toward saturation. As a result of the difference in flux levels in the two cores, core IV reaches saturation sooner than core I and a step of voltage, $e_o = -e_{s2}$, appears across the output transformer. When core I saturates, the net output voltage drops to zero. The same sequence of events occurs for cores II and III on the next half-cycle, and the net output is therefore a series of pulses at a repetition rate of 10 kc. The polarity of the output pulses is determined by the polarity of the control signal, and no out-

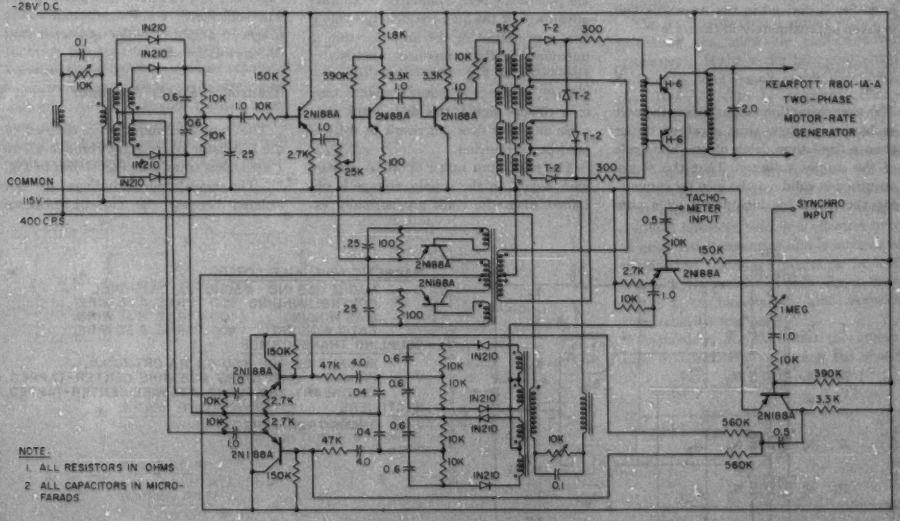


Fig. 2. Electric circuit diagram

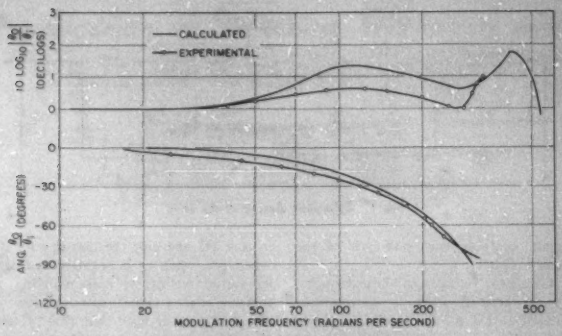


Fig. 3. Frequency response of the servo

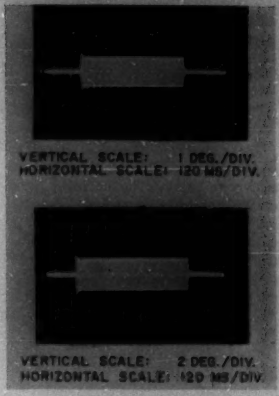


Fig. 4. Transient response of the servo

put occurs for a zero control signal because both cores saturate simultaneously.

The operation described gives rise to a dual-carrier system consisting of the system 400-cps carrier signal and the 10-kc repetition frequency of the output pulses. Both the 400-cps carrier and the error signal acting as amplitude modulation upon it act as modulation to vary the width and polarity of the high-frequency output pulses from the magnetic amplifier. Fig. 6 illustrates the ideal magnetic-amplifier output for a 400-cps signal input.

The transfer function for the magnetic amplifier is derived in the Appendix and is given approximately in the form

$$G_{MA}(S) = K_{MA}\frac{1}{\tau_{MA}S+1} \qquad (1)$$

This approximation characterizes the unit as having a single time constant rather than a fixed time delay of one half-cycle of the supply voltage. That this representation is valid may be seen by computing the phase lag introduced by a pure transportation delay at a signal frequency of 300 radians per second:

$$\frac{\pi\times 300}{2\pi\times 5{,}000}\times 57.2 = 1.7°$$

By way of comparison, the phase lag at a signal frequency of 300 radians per second of a similar unit utilizing a 400-cps carrier would be

$$\frac{\pi\times 300}{2\pi\times 400}\times 57.2 = 21.5°$$

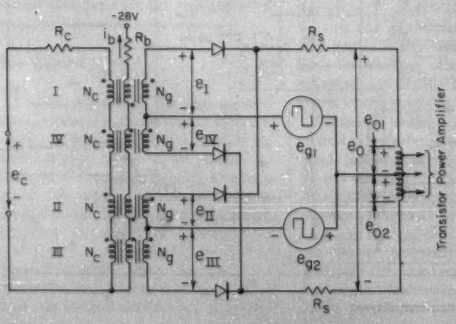
These figures show that the objectionable time delay of the magnetic amplifier has practically been eliminated. Another advantage of the high-frequency supply is that any output of the amplifier due to unbalance in the components consists of harmonics of the 10-kc repetition frequency. Since the motor responds only to a 400-cps signal, this drift in the magnetic amplifier does not result in drift of the output position.

The maximum range of linear control with regard to the supply voltage is obtained when the cores are normally excited. (Normal excitation is defined as the supply voltage amplitude necessary to change the core flux from minus to plus saturation in a period of one half-cycle.)

Because the supply voltage is usually set for normal excitation of the cores, the peak current in the output circuit may be adjusted for the desired value by the series resistor R_s. The peak current is determined by the base drive necessary to switch the transistors for the largest load expected. In this case, 200 milliamperes was sufficient. Total dissipation in the magnetic amplifier under quiescent conditions was 2 watts.

Transistor Power Amplifier

The magnetic amplifier is coupled to the transistor power stage through a 3 to 1 stepdown transformer as shown in Fig. 7. The transformer windings are not isolated because the current-resistance



MAGNETIC AMPLIFIER CORERS
MAGNETICS INC. #5000-2A ORTHONOL
CONTROL WINDING 100 TURNS #36 WIRE
BIAS WINDING 100 TURNS #36 WIRE
GATE WINDING 400 TURNS #36 WIRE
COUPLING TRANSFORMER
MAGNETICS INC. #50168-2A ORTHONOL
PRIMARY 840 TURNS #35 WIRE CENTER-TAPPED
SECONDARY 280 TURNS #35 WIRE CENTER-TAPPED
RECTIFIERS
TRANSITRON TYPE T-2

Fig. 5. Magnetic amplifier

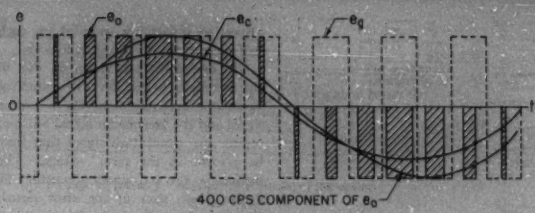


Fig. 6. Magnetic amplifier output for a sinusoidal input

drop in the windings due to the exciting current and the quiescent current in the magnetic-amplifier output circuit provide a positive bias on the transistor bases. This positive bias, for the p-n-p transistor, serves to prevent current flow from emitter to collector under zero-signal conditions.

When a step of voltage appears across the coupling transformer $e_o = e_{o1}$, the base transistor A is pulsed positive with respect to its emitter, and the base of transistor B is pulsed negative with respect to its emitter. If sufficient base drive is provided by the magnetic amplifier, the voltage across the emitter-collector junction of transistor B will drop to almost zero while transistor A will remain cut off. Transistor B is again switched off when the net magnetic-amplifier output drops to zero.

Analysis of the power stage is complicated by the tuned network load. In fact, it is not possible to derive a transfer function for the output amplifier alone because the output impedance changes radically as the transistors are switched on and off. An approximate transfer function from magnetic-amplifier input to power amplifier output which checks well with experimental data for small signals may, however, be derived from the following considerations.

The transistor power amplifier operates in one of the two states: 1. both transistors nonconducting, and 2. one transistor conducting and the other transistor nonconducting. In state 1, the output impedance is high, whereas in state 2 the output impedance is low. Because the ON time is proportional to the signal amplitude, the amplifier impedance is predominantly high for small-loop error signals. Furthermore, the transistors are current-limited, so that for small signals the load is driven by a current pulse of roughly constant amplitude. In this case the amplifier may be represented as a current source. Such an assumption represents the worst possible case from a dynamic standpoint.

The transfer function given in equation 2 follows from the foregoing discussion.

$$\frac{e_{m(s)}}{e_{b(s)}} = K_M G_{MA}(S) K_{PA}(S) \qquad (2)$$

$K_{PA}(S)$ is determined from the impedance loading the assumed output current source and is of the form

$$K_{PA}(S) = \frac{1}{TS+1}$$

However, the tuned motor represents a low Q tank circuit so that the following approximation is valid over a fairly large range of signal frequencies:

$$K_{PA}(S) \approx e^{-\tau s}$$

where $L_m/R_m \leq \tau \leq 2L_m/R_m$ and L_m = inductance of the motor control windings, and R_m = equivalent stalled-rotor-resistance. This approximation provides a conservative prediction of the forward loop phase lag, and is therefore used in the analysis.

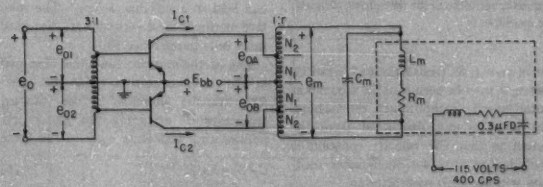
Compensation

To improve the performance of the servo, three basic types of compensation were initially considered: 1. viscous-coupled inertia damper, 2. lead network, and 3. tachometer feedback.[8]

Equalization by means of an inertia damper coupled to the load has the advantage that both high velocity and torque constants are obtainable, plus a wide bandwidth.[8] The damper, however, draws power from the load, under transient conditions, which has two degrading effects. First, it decreases acceleration capabilities of the motor and, second, it produces a long, nonlinear oscillating transient following a large step input. By declutching the damper during slew, these difficulties can be overcome. The complexity of such a system, however, prompted investigation of a simpler solution.

The lead network is a simple means of increasing the allowable gain-crossover frequency. If the motor time constant is sufficiently long (small motor damping), the allowable velocity constant is also improved over the uncompensated system. The torque constant for the system employing lead network equalization is low, but this may be improved by the use of an integral network. The integral network increases only the "static" torque constant however, since it is effective only at quite low frequencies and is ineffective in offsetting dynamic changes in the load torque. To avoid this difficulty, a combination of tachometer and network compensation was used.

Tachometer feedback has the advantage of effectively increasing the motor



OUTPUT TRANSFORMER
ARNOLD ENGINEERING CO. #5233-D2 DELTAMAX
N_1 = 200 TURNS #30 WIRE
N_2 = 400 TURNS #30 WIRE
TRANSISTORS
HONEYWELL H-6
SUPPLY VOLTAGE
28 VOLTS DC

Fig. 7. Transistor power amplifier

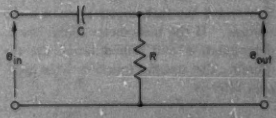


Fig. 8. High-pass filter

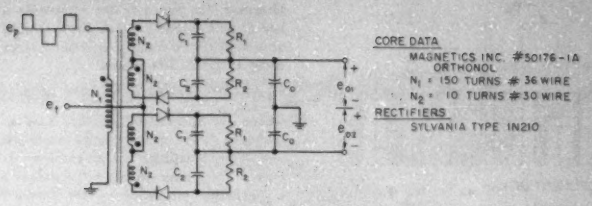


Fig. 9. Magnetic demodulator

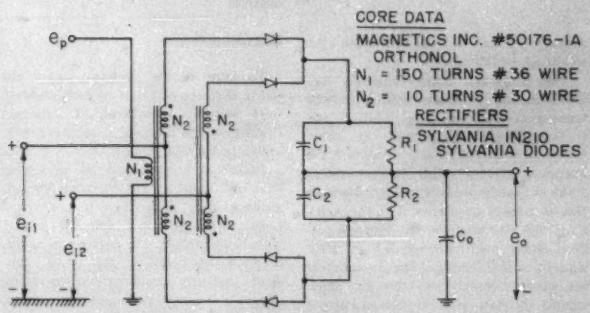


Fig. 10. Magnetic modulator

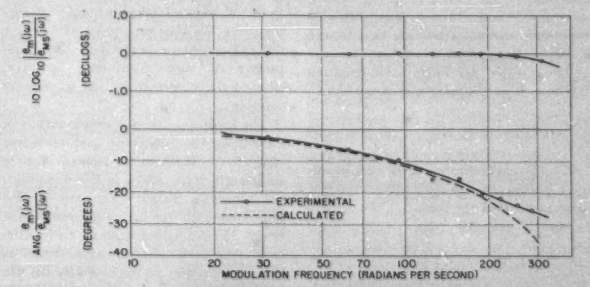


Fig. 11. Comparison of calculated and measured frequency response of the modulator, magnetic amplifier, and power amplifier

damping without drawing power from the load. As a result, the time constant of the motor is effectively reduced so that the open-loop gain and the bandwidth of the system may be increased. The low-frequency gain may be further increased by passing the tachometer signal through the simple high-pass filter shown in Fig. 8 provided the tachometer signal is a d-c voltage. If the tachometer signal is a-c, demodulation and remodulation are required. The transfer function of the high-pass filter is

$$\frac{e_{out}(S)}{e_{in}(S)} = \frac{S}{S+\omega_f} \qquad (3)$$

where $\omega_f = 1/RC$(resistance-capacitance). Hence, it may be seen that for a constant motor speed (ramp input), the tachometer is effectively decoupled from the system so that the velocity constant is much higher than for tachometer feedback without high-pass filtering

Modulator and Demodulator

The magnetic-keyed, phase-sensitive sampling units used to demodulate and remodulate the tachometer output are basically identical to the device described by Mark, et al.[9] To reduce the phase lags introduced by these units, their respective inputs are sampled on each half-cycle of the 400-cps system carrier.

The demodulator is shown in Fig. 9; the modulator in Fig. 10. The sampling modulator not only sums the a-c error signal and d-c tachometer signal but also rejects any quadrature signal that may be present in the error voltage. The modulator also introduces a constant 90 degree phase shift of the error signal. These effects are the result of the sampling being performed at the peak of the 400-cps reference voltage wave.

The transfer functions of the demodulator and of the modulator can be approximated by a simple time delay of 1/4 cycle of the system carrier frequency, and are therefore of the form

$$G_{DS} = K_{DS} e^{-(\pi/2\omega_c)S} \qquad (4)$$

$$G_{MS} = K_{MS} e^{-(\pi/2\omega_c)S} \qquad (5)$$

Both K_{DS} and K_{MS} are dependent upon the output capacitance C_o and the keying pulse amplitude. Selection of these values for this particular application is discussed in reference 9.

Analysis of the Servo

As previously shown, the time delay of the magnetic amplifier using a 5-kc supply is negligible. The effects of the amplifier time constant τ_{MA} may also be made negligible by proper adjustment of the control resistance.

For small R_c value, amplifier gain is high, but the time constant also assumes a large value. If τ_{MA} is made suitably large, however, the modulating signals are essentially shifted by a constant 90 degrees, and this phase lag may be cancelled by shifting the reference phase of the motor by 90 degrees. For a large value of R_c, both the amplifier gain and time constant are small, so that either extreme can lead to a suitable design. The relative merits of these extremes are discussed more fully in reference 10.

Use of the transfer functions obtained in the previous section yields the open-loop transfer function of the forward transmission (excluding the motor dynamics and neglecting τ_{MA}) in the form

$$G_f(S) = K_\epsilon e^{-(\pi/2\omega_c + t)S} \qquad (6)$$

The validity of this approximation is confirmed by the experimental data plotted in Fig. 11.

The effect of the modulator, aside from the constant attenuation K_{MS}, is to introduce a time delay of 1/4 cycle of the carrier frequency in the position loop. Because the modulator and demodulator sample at almost the same

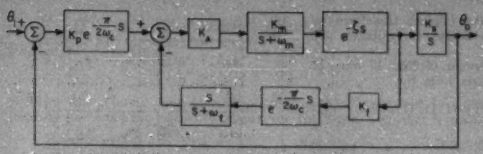


Fig. 12 (above). Mathematical model of the system

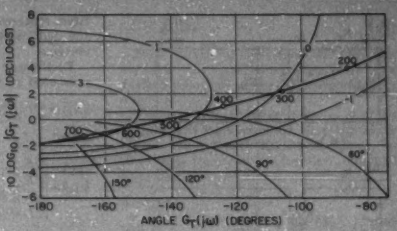


Fig. 13 (right). Nichols plot for the tachometer loop

time, the modulator introduces a negligible time delay of the tachometer signal in addition to that already introduced by the demodulator. Summation of position and acceleration error is therefore effectively performed at the modulator output terminals except for the demodulator attenuation. With K_{MS} lumped into K_A and the transportation lag of the modulator acting only upon the position error, the system may be represented by the block diagram of Fig. 12.

The open-loop transfer function for this mathematical model is

$$G(S)=(K_pK_A/K_f)\frac{S+\omega_f}{S^2}W_T(S) \qquad (7)$$

where $W_T(S)$ is the closed-loop weighting function of the tachometer loop, given by

$$W_T(S)=\frac{G_T(S)}{1+G_T(S)} \qquad (8)$$

and

$$G_T(S)=K_AK_mK_f\frac{e^{-(\pi/2\omega_c+\tau)S}}{S+\omega_f} \qquad (9)$$

For the type of motor driving stage used in this system, the motor damping is extremely small for low-shaft velocities. Hence, $\omega_m \approx 0$, and

$$G_T(S)\approx K_AK_mK_f\frac{e^{-(\pi/2\omega_c+\tau)S}}{S+\omega_f} \qquad (10)$$

A Nichols chart plot for the tachometer loop is shown in Fig. 13 for $\omega_f=100$ radians per second. The open-loop transfer function, obtained from this plot and equation 7, is shown in Fig. 14. The effect of the time delay present in the power stage and in the demodulator is apparent from Fig. 13, which shows the low rate of attenuation and rapid phase shift with increasing frequency. The tachometer loop was adjusted for a resonant frequency of 700 radians per second with a peak closed-loop magnitude ratio of 3.8 decibels.

The effect of the narrow resonant peak of $W_T(S)$ on the system open-loop transfer function may be seen in Fig. 14 by the deviation of the actual magnitude from that of the logarithmic asymptotes. The rapid phase shift with almost no attenuation at frequencies between 300 and 700 radians per second requires that the gain crossover be made very near the corner frequency of the high-pass filter.

Fig. 1 shows a comparison of the calculated and the experimental closed-loop frequency response of the system. The experimentally determined magnitude ratio is seen to be somewhat less than the predicted value. Although equipment was not available for obtaining signal frequencies above 314 radians per second, comparison of the results indicates that the resonant frequency is between 350 and 400 radians per second and that the peak magnitude ratio M_p is between 1.5 and 2.0.

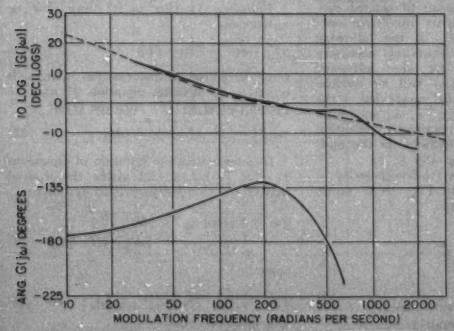


Fig. 14. Attenuation and phase diagram for the mathematical model

Appendix

The following derivations are based upon a new method of magnetic-amplifier analysis proposed by Johannessen.[11] The procedure is based upon representation of the half-cycle averages of the terminal variables of a basic single-core amplifier, shown in Fig. 15, in matrix form. (In Fig. 15, the core is represented by an ideal reactor shunted by a linear magnetizing resistance R_M. This assumption is discussed in detail in reference 11, pages 9–15, and experimental confirmation of its validity has been obtained by Knutrud.[12])

Once the terminal variables are established, this basic element may be interconnected with other basic circuits to yield the multicore amplifier configuration under consideration. The difference equation[13] for the multicore amplifier is then obtained from matrix manipulations.

The 4-core amplifier used in the system is redrawn in Fig. 16 for convenience in

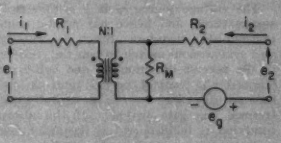


Fig. 15. Basic amplifier circuit

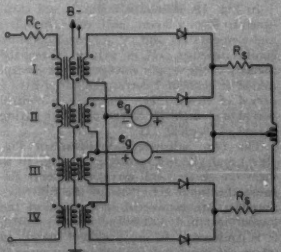


Fig. 16. Full-wave amplifier

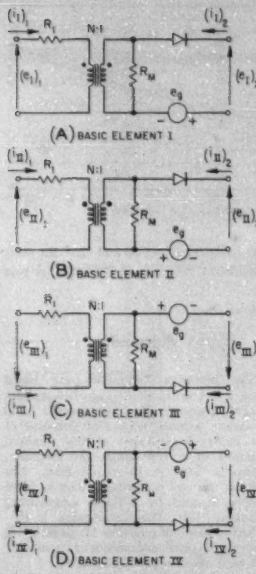


Fig. 17. Basic elements of the full-wave amplifier

separating the device into four basic elements. These four single-reactor elements are shown separately in Fig. 17 in order to define the reference polarities of the circuit variables.

The voltage source e_g is a symmetrical square wave, and hence has no d-c component. It therefore functions merely as a synchronous switch to bias the rectifier in the reverse direction on one half-cycle (termed the control period) and in the forward direction on the next half-cycle (termed the gate period). If the above condition is fulfilled and if the amplitude and the frequency of e_g do not fluctuate, then the supply voltage does not enter into the relations between the terminal variables.

In Fig. 18, elements I and II are combined to form a 2-core, half-wave amplifier, and elements III and IV are combined in a similar manner. It is evident from the polarity of the supply voltages that a control period of element I corresponds to a gating period of element II. Similarly, a control period of element IV corresponds to a gating period of element III.

The two half-wave amplifiers of Fig. 18(A) and (B) are combined in Fig. 19 to form the desired full-wave amplifier. Because the input terminals of the full-wave amplifier are connected in series and the output terminals are connected in parallel, it is convenient to use a hybrid matrix in which the input voltage and output current are used as the dependent variables. In the following equations, the variable quantities represent the Laplace transforms of the half-cycle averages of voltages and currents.

Each basic element of Fig. 17 may be presented in matrix form by

$$\|\alpha_j\| = \|H_j\| \ \|\beta_j\| \ (j = \mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IV}) \qquad (11)$$

where

$$\|\alpha_j\| = \begin{Vmatrix} (E_j)_1^c \\ (E_j)_1^g \\ (I_j)_2^c \\ (I_j)_2^g \end{Vmatrix} \text{ and } \|\beta_j\| = \begin{Vmatrix} (I_j)_1^c \\ (I_j)_1^g \\ (E_j)_2^c \\ (E_j)_2^g \end{Vmatrix} \qquad (12)$$

The superscripts c and g refer to the control and gating periods respectively.

The basic circuit of Fig. 15 shows the series resistor in the output circuit R_b as a linear element. This resistor must be modified to take into account the diode in the output circuit of the actual elements shown in Fig. 17. In the following equations, the term R_b is defined as the diode reverse resistance plus the resistance of the core winding, and R_f is defined as the diode forward resistance plus the resistance of the core winding. Thus, R_f replaces R_b when e_g is in such a direction that the rectifier is conducting, and R_b replaces R_b when e_g is in such a direction that the rectifier is nonconducting.

Because the basic elements are identical, the $\|H_j\|$ matrix is the same for any value of j. With the modifications explained in the above paragraph taken into account, the network matrix is given by

$$\|H_1\| = \begin{Vmatrix} (R_1 + N^2 R_b) & 0 & NR_b/R_b & 0 \\ -N^2 R_b' & R_1 & -NR_b'/R_b & 0 \\ -NR_b/R_b & 0 & R_b/R_M R_b & 0 \\ NR_b'/R_f & 0 & R_b'/R_f R_b & 1/R_f \end{Vmatrix} \qquad (13)$$

where

$R_b = R_M R_b/(R_M + R_b)$

$R_b' = R_b/\tau$

$\tau = \epsilon^{(\pi/\omega_0)\delta}$

ω_0 = frequency of the supply voltage e_g.

Because of the symmetry which exists between the interconnection consisting of elements I and II and the interconnection consisting of elements III and IV, it is apparent that for any equation involving only elements I and II, an identical equation (except for a change of capital subscripts) may be written involving only elements III and IV. In the following steps, therefore, only equations necessary to obtain relations for the half-wave amplifier of Fig. 18(A) will be written explicitly. The terminal quantities of the final interconnection (e_1, i_1, e_2, and i_2 in Fig. 19) will then be obtained by addition of the two identical matrices for the half-wave amplifiers of Fig. 18.

Consideration of the relative polarity of the supply voltages in Fig. 18(A) yields the following relations:

$$\begin{aligned} (I_1)_1^c &= (I_{11})_1^g \\ (I_1)_1^g &= (I_{11})_1^c \\ (E_1)_2^c &= (E_{11})_2^g \\ (E_1)_2^g &= (E_{11})_2^c \end{aligned} \qquad (14)$$

If the transformation matrix

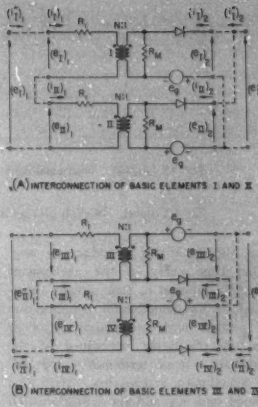


Fig. 18. Interconnections to form two half-wave amplifiers

$$\|T\| = \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{Vmatrix} \qquad (15)$$

is defined, it may be seen that

$$\|\beta_{\mathrm{II}}\| = \|T\| \ \|\beta_{\mathrm{I}}\| \qquad (16)$$

From Fig. 18(A), the following relations may be written:

$$(E_1'')_1 = (E_{\mathrm{I}})_1^c + (E_{\mathrm{II}})_1^g \qquad (17)$$

$$(I_1'')_2 = (I_{\mathrm{I}})_2^g + (I_{\mathrm{II}})_2^g \qquad (18)$$

The following transformation of variables is now made:

$$\|\alpha_{\mathrm{II}}\| = \|T\| \ \|\alpha_{\mathrm{II}}'\| \qquad (19)$$

where

$$\|\alpha_{\mathrm{II}}'\| = \begin{Vmatrix} (E_{\mathrm{II}})_1^g \\ (E_{\mathrm{II}})_1^c \\ (I_{\mathrm{II}})_2^g \\ (I_{\mathrm{II}})_2^c \end{Vmatrix} \qquad (20)$$

Equation 11 (for $j = \mathrm{II}$) may now be written as

$$\|T\| \ \|\alpha_{\mathrm{II}}'\| = \|H_{\mathrm{II}}\| \ \|T\| \ \|\beta_{\mathrm{I}}\| \qquad (21)$$

but, since the transformation matrix is equal to its inverse, equation 21 may be premultiplied by $\|T\|$ with the result

$$\|\alpha_{\mathrm{II}}'\| = \|T\| \ \|H_{\mathrm{II}}\| \ \|T\| \ \|\beta_{\mathrm{I}}\| \qquad (22)$$

It is seen that the addition of equations 11 (for $j = \mathrm{I}$) and 22 yields the desired matrix for the double-primed terminals.

$$\|\alpha_{\mathrm{I}}''\| = \|H_{\mathrm{I}}\| + \|T\| \ \|H_{\mathrm{II}}\| \ \|T\| \|\beta_{\mathrm{I}}\| = \|H\| \ \|\beta_{\mathrm{I}}\| \qquad (23)$$

where

$$\|H\| = \|H_{\mathrm{I}}\| + \|T\| \ \|H_{\mathrm{II}}\| \ \|T\| \qquad (24)$$

and

$$\|\alpha_I''\| = \|\alpha_I\| + \|\alpha_I'\| = \begin{vmatrix} (E_{I1})_1^c + (E_{II})_1^d \\ (E_{I1})_1^d + (E_{II})_1^c \\ (I_1)_3^c + (I_{II})_3^d \\ (I_1)_3^d + (I_{II})_3^c \end{vmatrix} = \begin{vmatrix} (E_1'')_1^c \\ (E_1'')_1^d \\ (I_1'')_3^c \\ (I_1'')_3^d \end{vmatrix}$$

When the matrix transformation and addition indicated in equation 24 is performed, the following network matrix for the 2-core amplifier is obtained as

$$\|H\| = \begin{vmatrix} (2R_1+N^2R_3) & -N^2R_3' & NR_3/R_5 & -NR_3'/R_5 \\ -N^2R_3' & (2R_2+N^2R_4) & -NR_3'/R_5 & NR_3/R_5 \\ -NR_3/R_5 & NR_3'/R_f & (1/R_f+R_3/R_MR_5) & R_3'/R_fR_5 \\ NR_3'/R_f & -NR_3/R_5 & R_3'/R_fR_5 & (1/R_f+R_3/R_MR_5) \end{vmatrix} \quad (25)$$

If the first two rows and the second two rows respectively of $[\alpha_I'']$ are added, there results

$$(E_1'')_1 = m(I_1'')_1 + n(E_1'')_3 \quad (26)$$

$$(I_1'')_3 = p(I_1'')_1 + q(E_1'')_3 \quad (27)$$

where

$$(E_k'')_r = (E_k'')_r^c + (E_k'')_r^d \quad (k = I, II)$$

$$(I_k'')_r = (I_k'')_r^c + (I_k'')_r^d \quad (r = I, 2)$$

$$m = 2R_1 + N^2(R_3 - R_3')$$

$$n = N/R_5(R_3 - R_3')$$

$$p = N(R_3'/R_f - R_3/R_5)$$

$$q = 1/R_f + R_3/R_MR_5 + R_3'/R_fR_5$$

It should be remembered at this point that an identical set of equations exists for the variables $(E_{II})_1$ and $(I_{II}'')_3$. From Fig. 18, it may be seen that

$$(I_{II}'')_1 = -(I_1'')_1 \quad (28)$$

and from Fig. 19 that

$$(I_1'')_1 = I_1 \quad (29)$$

Hence, the network matrices for the double-primed terminals of the half-wave circuits of Fig. 18(A) and (B) may be written as

$$\begin{vmatrix} (E_1'')_1 \\ (I_1'')_3 \end{vmatrix} = \begin{vmatrix} m & n \\ p & q \end{vmatrix} \begin{vmatrix} (I_1'')_1 \\ (E_1'')_3 \end{vmatrix} \quad (30)$$

$$\begin{vmatrix} (E_{II}'')_1 \\ (I_{II}'')_3 \end{vmatrix} = \begin{vmatrix} -m & n \\ -p & q \end{vmatrix} \begin{vmatrix} (I_1'')_1 \\ (E_{II}'')_3 \end{vmatrix} \quad (31)$$

Consideration of Fig. 20 leads to the following equations:

$$(E_1'')_3 = -(I_1'')_3R_s + E_2 \quad (32)$$

$$(E_{II}'')_3 = -(I_{II}'')_3R_s - E_2 \quad (33)$$

Substitution of equation 32 into 30 and equation 33 into 31 yields

$$\begin{vmatrix} (E_1'')_1 \\ (I_1'')_3 \end{vmatrix} = \begin{vmatrix} A & B \\ C & D \end{vmatrix} \begin{vmatrix} I_1 \\ E_2 \end{vmatrix} \quad (34)$$

$$\begin{vmatrix} E_{II}'')_1 \\ I_{II}'')_3 \end{vmatrix} = - \begin{vmatrix} A & B \\ C & D \end{vmatrix} \begin{vmatrix} I_1 \\ E_2 \end{vmatrix} \quad (35)$$

where

$$A = m - \frac{npR_s}{1+qR_s}$$

$$B = n - \frac{nqR_s}{1+qR_s}$$

$$C = \frac{p}{1+qR_s}$$

$$D = \frac{q}{1+qR_s}$$

Subtraction of equation 35 from equation 34 yields the network matrix for the full-wave amplifier:

$$\begin{vmatrix} E_1 \\ I_2 \end{vmatrix} = 2 \begin{vmatrix} A & B \\ C & D \end{vmatrix} \begin{vmatrix} I_1 \\ E_2 \end{vmatrix} \quad (36)$$

$$E_1 = (E_1'')_1 - (E_{II}'')_1$$

$$I_2 = (I_1'')_3 - (I_{II}'')_3$$

To relate the output voltage to the control voltage, the load characteristics must be known. If it is assumed that the input impedance to the following stage is a linear resistance R_T, then

$$I_2 = -E_2/R_T \quad (37)$$

Substitution of equation 37 into the network matrix and elimination of I_1 from the two equations yields the desired transfer function.

The resulting transfer function may be simplified if it is assumed that the diode reverse impedance R_s is infinite. In this case, the transfer function obtained from equations 36 and 37 becomes

$$E_2(S) - 2\sigma\epsilon^{-(\pi/\omega_0)S}E_2(S) = -(2\gamma\sigma/N)\epsilon^{-(\pi/\omega_0)S}E_1(S) \quad (38)$$

or

$$\frac{E_2(S)}{E_1(S)} = -\frac{2\gamma\sigma}{N} \frac{\epsilon^{-(\pi/\omega_0)S}}{1-2\sigma\epsilon^{-(\pi/\omega_0)S}} \quad (39)$$

where

$$\sigma = \frac{N^2R_M}{R_c + 2N^2R_M}$$

$$R_c = 4R_1$$

$$\gamma = \frac{R_T}{R_f + R_3 + 2R_T}$$

The difference equation may be obtained by taking the inverse Laplace transform of equation 38. The result of this operation is

$$E_2(n+1) - 2\sigma E_2(n) = -(2\gamma\sigma/N)E_1(n) \quad (40)$$

Equation 39 is somewhat inconvenient for the purpose of servo analysis because of the delay term in the denominator. Since $\pi/\omega_0 \ll 1$, the exponential term may be approximated by the first two terms of a series expansion:

$$\epsilon^{-(\pi/\omega_0)S} \approx \frac{1}{(\pi/\omega_0)S+1}$$

Equation 39 then becomes

$$G_{MA}(S) = \frac{E_2(S)}{E_1(S)} = -\frac{2\gamma\sigma}{N(1-2\sigma)} \frac{1}{\frac{(\pi/\omega_0)}{1-2\sigma}S+1} \quad (41)$$

Equation 41 represents the amplifier by a single time constant and neglects the inherent transportation lag of one half-cycle of the supply voltage, which is assumed to be negligible.

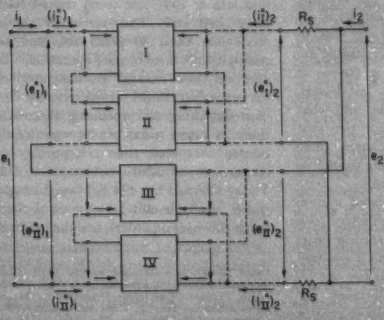


Fig. 19. Full-wave amplifier as an interconnection of four basic elements

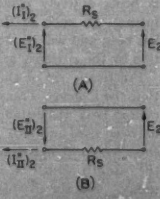


Fig. 20. Relationships at the output terminals

References

1. Application of a Magnetic Amplifier to a High Performance Instrument Servo. P. R. Johannessen. *Technical Report 7002-10*, Massachusetts Institute of Technology, Cambridge, Mass., May 1955.

2. An Improved Servo Amplifier. C. W. Lufcy, A. E. Schmidt, P. W. Barnhardt. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, Sept. 1952, pp. 281–89.

3. A Five-Kilocycle Magnetic Amplifier for Servo-Motors, D. O. Fischer. *M.S. Thesis*, Massachusetts Institute of Technology, Aug. 1956.

4. Power Switching Transistors in Servo Amplifiers, S. E. Beckjord. *Technical Report 7002-13*, Massachusetts Institute of Technology, June 1956.

5. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H.

Royer. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, July 1955, pp. 323–33.

6. Comparison of Three Basic Types of Compensation of Electric Servomechanisms, George Biernson. *Engineering Memorandum* no. 27, *DIC Project 6506*, Massachusetts Institute of Technology, May 1952.

7. Damper Stabilized Servo Data Repeater, J. E. Ward. *Report 7138-TM-3*, Massachusetts Institute of Technology, July 1955.

8. A Study of a Dual-Mode Damper-Stabilized Servo, James R. Kaiser. *Technical Report 7002-16*, Massachusetts Institute of Technology, June 1956.

9. Magnetically Keyed, Phase-Sensitive Demodulators, Robert B. Mark, Wheeler X. Johnson, Paul R. Johannessen. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Mar. 1957, pp. 1–6.

10. Application of Switching Transistors and Saturable Reactors in a High-Performance Servo, Fred B. Cox, Jr. *M.S. Thesis*, Massachusetts Institute of Technology, Sept. 1957.

11. Analysis of Magnetic Amplifiers, P. R. Johannessen. *Memorandum 7668-M-235 R*, Massachusetts Institute of Technology, July 2, 1957.

12. Analysis and Design Procedure for a Two-Stage Magnetic Amplifier, T. Knutrud. *M.S. Thesis*, Massachusetts Institute of Technology, June 1956.

13. Analysis of Magnetic Amplifiers by the Use of Difference Equations, P. R. Johannessen. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 700–11.

Rapid Transit Is Expanding in Chicago

STANLEY D. FORSYTHE
NONMEMBER AIEE

MAN is that inexplicably complex creature who confidently expects to fly to the moon in the next decade and yet spends more time between home and work now than he did a decade ago.

When the mayors of some of the world's largest cities were asked by the Citizens' Budget Commission of New York, N.Y., to name the worst problems facing their cities, finance, taxation, and transportation and traffic were almost unanimously selected.[1]

A Chicago, Ill., planner has estimated that it takes 5,000,000 square feet of off-street parking space to handle the automotive traffic brought in by a single new expressway lane, implying that the mammoth road building programs under way will help provide facilities to carry vehicles in motion, but will not solve the problem of providing for these vehicles at rest.

Every city in the United States underwrites its rapid transit system to some degree, and those now without rapid transit but planning such facilities for the future, will underwrite these operations to a greater degree than most of the existing systems enjoy.

There would seem to be a common denominator to these rather unrelated statements. It is this: A modern well-designed, adequate rapid transit system can contribute substantially to solving the mayors' problems, can contribute substantially to minimizing the need for mammoth downtown parking facilities, and can, in general, improve the attractiveness of urban living. More and more it is being recognized throughout the United States that a rapid transit system is a very valuable municipal asset which should be considered in much the same light as are other municipal assets such as highways, sewers, water, lighting, fire protection, and police protection. In San Francisco, Calif., for example, legislation is being prepared which would provide the necessary measure of public support for a rapid transit system that is estimated will cost ultimately $1,700,000-000. It will serve a population estimated at 7,000,000 by 1990 and a territory of approximately 7,500 square miles. A somewhat similar program is under way in Los Angeles and a number of other cities throughout the country are taking tentative steps in this direction. Therefore, since every rapid transit system on the North American continent in operation or under consideration is supported or planned to be supported to a greater or less degree by the municipality which it serves, it is not too radical to say that any large city can have just as fine a rapid transit system as its citizens desire.

Chicago planners and government officials are awake to this possibility and the city is becoming increasingly aware of the value of its rapid transit system. As this awareness grows, another decade should see tremendous changes and significant enlargement of the existing system which, during the last 40 of its 55 years of private operation, was financially unable to make any sizable extensions. This prediction is based on what has been done during the last decade, what is being done at present, and what is being actively planned for the near future. In each of these phases, there has been and will be increasing participation by the citizenry as a whole.

Practically all of the elevated system that the Chicago Transit Authority took over in 1947 was installed in the few years between 1893 and 1907. That it was well conceived and well built is borne out by the use to which it is still being put today. But the city has grown so greatly both in population and area that rapid transit routes in some important instances do not extend far enough to serve sizable residential, commercial, and manufacturing centers, and in other instances are nonexistent.

The cost of underground tubes is prohibitively expensive and they must be confined to those areas where no other type of right-of-way is available. The outmoded elevated structures erected around the turn of the century depress property values and discourage substantial additions to this type of right-of-way. There remains, however, the unparalleled opportunity to incorporate rapid transit into urban highway plans, and this possibility is being actively considered in connection with every expressway which is planned for the Chicago area. In addition, there is gradually emerging a realization that railroad rights-of-way furnish well-protected traffic-free arteries that may also, in certain instances, accommodate rapid transit facilities at great benefit to the municipality. A look at a railroad map of any large city will show how much comparatively lightly used right-of-way of this type exists. The rapidly changing attitude of railroad officials towards this possibility makes it a particularly attractive one to explore when rapid transit routes are being sought.

When the Chicago Transit Authority took over the existing rapid transit system late in 1947, the rapid transit company had only four pieces of equipment less than 30 years old, and no multiple-unit door-control operation. In addition, there were far too many stations for true rapid transit service, and the cost of maintaining and operating these stations 24 hours a day was a very heavy burden. At that time the number of employees per vehicle amounted to 3.2. Today Chicago has 474 lightweight high-speed multiple-unit door-control cars (with 130 more on order), and the multiple

Paper 56-837, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department. Manuscript submitted June 26, 1956; made available for printing April 3, 1959.

Stanley D. Forsythe is with the Chicago Transit Authority, Chicago, Ill.

unit door control has been installed on more than 400 of the older all-steel cars. Approximately 25 stations have been eliminated, fare collection on the trains has been instituted during the hours of lighter traffic, and several terminal loops have been installed to replace stub-end operations with their attendant operating costs. The skip-stop system in wide use on the entire system has substantially increased the operating speed of all trains. The result of this and other kindred changes has been to reduce the number of employees per vehicle to 2.4. All this has been accomplished with but one significant change in the existing routes: the opening of the Dearborn Street Subway. This change provided a more direct downtown delivery for the Logan Square Branch extending approximately 5 miles northwest from the Loop.

All urban transportation systems have seen drastic changes in revenue passengers during the last decade. The Chicago Transit Authority was no exception. However, the way in which rapid transit riding continued, and even increased slightly, during this period greatly encouraged those who feel that rapid transit is an important asset to Chicago's well-being. While revenue ridges on the surface system declined from 641,600,000 in 1950 to 510,600,000 in 1955 (a decrease of approximately 20%), the rapid transit riding increased from 110,600,000 to 112,900,000. This is a very modest increase, but a very significant one in view of the fact that there was no change in the territory directly served by the rapid transit system during these years.

During this 6-year period the car-miles per employee on the entire system increased from 8,750 to 10,800, or 23%. The corresponding increase on the Rapid Transit Division was from 9,150 to 12,250, or 34%.

The revenue rides per employee on the entire system increased from 38,300 to 41,500, or 10.5%. The corresponding increase on the Rapid Transit Division was from 25,200 to 31,800, or 26%.

These figures are measures of operating efficiencies and illustrate what is being done in rapid transit service to reduce costs with modern facilities and techniques.

Much more can be done and is being planned in this area. A further and more complete modernization of rolling stock, substitution of high-speed grade-separated private rights-of-way for obsolete elevated structures and ground-line operation, simplified fare-collection techniques, modern maintenance facilities, and other related activities will tend to decrease the operating costs of rapid transit operation further.

It was stated earlier that any large city have as fine a rapid transit system as its citizens desire.

The preceding figures indicate that even today, modern rapid transit operation can be in the same cost bracket as bus operation on city streets, but that the capital cost remains a burden. This suggests that cities and rapid transit operators might co-operate to their mutual advantage on a "we'll build it, you operate it" basis. The value that the San Francisco planners place on this type of public works is clearly indicated by the sources of funds suggested to finance their $1.7 billion rapid transit program, the first stage of which is estimated to cost $875,000,000. They feel that the need for rapid transit in this area is such that they would provide over $30,000,000 of public support per year for the next 30 years to serve the bonded indebtedness of the first stage. This public support would come from property tax, retail sales tax, gasoline tax, and payroll tax. In addition, a proportion of bridge tolls is included. The possibility of state or federal aid has been discussed, but the planners prudently refrain from recommending a definite proportion of support from these sources.

It is clearly intended, however, that the citizens acting through either an authority or a district (probably the latter in this instance) will finance and build the rapid transit facilities and that the operation of these facilities will be supported by the revenues. That degree of support could revolutionize the activities of many large American urban areas, and the planners are to be commended for their vision.

The rapid transit incorporated in the median strip of the Congress Street or West Route Expressway, opened in June 1958, utilizes welded rails on heavily ballasted track with long radius curves to permit a speed of operation limited only by the distance between stops. Attractive island-type stations are connected by covered ramps to adjacent bridges carrying cross-traffic over the expressway. This route, extending approximately 10 miles west from the Loop, largely replaces an older route that combined elevated and ground-line operation using some of the oldest equipment on the system. It has a new downtown delivery which, instead of using the elevated Loop, combines with the Logan Square Branch and operates through the Dearborn Street Subway. It will be extremely interesting to see the effect on traffic volume of new equipment, new right-of-way, and a greatly reduced running time.

The Northwest Route Expressway, extending from a point just west of the Loop to the northwest city limits, was originally designed without provision for rapid transit. In the last several years, however, the highway designers have become convinced that rapid transit in the median strip is a vital necessity in order to relieve the expressway of some of its automobile traffic. Therefore, the plans were altered to include a 2-track rapid transit line with island platform stations approximately every 7/10ths of a mile. This line extends northwest from the Logan Square terminal for a distance of approximately $7^1/_2$ miles. It was installed in accordance with the formula that has been used by the city and the Chicago Transit Authority since the inception of the Authority whereby the city installs all of the fixed transportation equipment and the Chicago Transit Authority repays the city in monthly payments over a period of approximately 30 years without interest charges. The cost of the right-of-way and stations is not included in the amount repaid to the city.

As the last three miles of this extension provide extremely light traffic for normal rapid transit operation, the Authority suggested to the city that the Authority be relieved of any repayment for the cost of fixed transportation equipment until the revenues justify such payments.

Another significant improvement in rapid transit service concerns the west end of the Lake Street Branch. The outermost $2^1/_2$-mile section of this line, which serves the west side of Chicago and the village of Oak Park, involves 22 grade crossings. Immediately to the north of the Chicago Transit Authority right-of-way is an elevated 6-track right-of-way belonging to the Chicago North Western Railroad. The railroad has indicated its willingness to lease two tracks to the Chicago Transit Authority and the various governmental and highway agencies interested in this $2^1/_2$-mile stretch have indicated their approval of the project and their willingness to share in the expense of eliminating these 22 grade crossings at an estimated total cost not to exceed $4,000,000. This would substantially improve the rapid transit service on this branch as well as immeasurably improve street traffic conditions through the affected area. There will also be a notable improvement in abutting property values. It is expected that work on this project can be completed 18 months after agreements are signed.

The South Route Expressway starts from a point just west of the Loop and, after a short swing to the east, generally parallels State Street south from the Loop to 99th Street where the highway divides. One branch will serve Beverly Hills and Morgan Park to the southwest while the other branch will serve the Lake Calumet industrial district to the southeast. Here, too, it is being planned that a median strip of sufficient width for rapid transit operation will be provided throughout the entire length of the expressway. It is felt that rapid transit operation should be provided in that part of the expressway nearest the central business district at the earliest opportunity and extended south as far and as quickly as circumstances permit. The south end of the State Street Subway would be only slightly more than $1^1/_2$ miles from a convenient point of access to this median strip, or a shorter connection could be made farther south to the south-side elevated structure.

Another opportunity for an important median strip rapid transit operation will arise when the the crosstown expressway is built. This is planned to extend 7 miles in a north-south direction, approximately $3^1/_2$ miles west of and paralleling State Street, from the Northwest Expressway on the north to the Southwest Expressway on the south. Rapid transit operation on this expressway which would intersect every rapid transit branch except the North-South and Ravenswood as well as 30-section line, half-section line, and diagonal streets, would furnish an excellent north-south trunk line by-passing the central business district but linking a number of outlying commercial, residential and manufacturing districts.

From the viewpoint of a mass transit operator, it is extremely gratifying to see how rapidly highway planners and city officials have come to realize the invaluable role that rapid transit plays in the big team job of moving people. As this realization grows and larger segments of cities' populations learn how much rapid transit can be provided for a comparatively little expenditure as compared with highway construction, one can expect to see greatly expanding rapid transit facilities in many American cities.

Reference

1. *American City*, New York, N. Y., Apr. 1958.

Analysis of A-C Servomotors Operated from Unbalanced Nonsinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances

E. R. LIND
ASSOCIATE MEMBER AIEE

N. L. SCHMITZ
MEMBER AIEE

MANY ANALYTICAL studies directed at the predetermination of output characteristics of a-c servomotors have been based on the assumptions of constant-source impedance and sinusoidal excitation. Under these assumptions, the method of symmetrical components serves admirably, particularly when steady-state characteristics are desired, and in cases where the duration of the mechanical transients makes the electrical transients insignificant.

In recent years, the use of devices such as chopper modulators, magnetic amplifiers, and gaseous tubes in servomechanism design has caused a-c servomotors to be subjected to new modes of operation. These devices may produce nonsinusoidal waveforms which are supplied through constant impedance amplifiers to the servomotor. In other cases the waveform is nonsinusoidal and the output impedance of the amplifier is nonlinear. With thyratrons or magnetic amplifiers, the output impedance is essentially zero during conduction and approaches open-circuit conditions during the nonconducting period. In addition, the waveform may have either an average value (d-c level), or the direct current may be superimposed on sine-wave excitation to produce some desired dynamic effect.

A considerable volume of literature has appeared concerning magnetic materials, thyratrons, semiconductor devices, etc. Much study has gone into these special devices but there is a serious lack of information available on the performance of a-c servomotors when operated in conjunction with these nonlinear devices.

Methods outlined in this paper circumvent the limitations of the method of symmetrical components and permit analysis of systems wherein source impedances are discontinuous functions of current and time. Electromotive force (emf) sources may be either sinusoidal or nonsinusoidal.

Paper 59-1157, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted March 20, 1959; made available for printing August 7, 1959.

E. R. LIND is with the General Electric Company, Ithaca, N. Y., and N. L. SCHMITZ is with the University of Wisconsin, Madison, Wis.

General Relationships

To specify completely the performance of a motor it is necessary to employ a physical representation in terms of the self and mutual operational impedances instead of the conventional steady-state equivalent circuit. This has been done in the classic paper by H. C. Stanley.[1] Stanley's equations can be rewritten in the following form to facilitate solution:

$$-\frac{dI_\alpha}{dt} = -\frac{V_\alpha}{L_s} + \frac{R_s}{L_s} I_\alpha + \frac{M}{L_s}\frac{di_\alpha}{dt}$$

$$-\frac{dI_\beta}{dt} = -\frac{V_\beta}{L_s} + \frac{R_s}{L_s} I_\beta + \frac{M}{L_s}\frac{di_\beta}{dt}$$

$$\frac{di_\alpha}{dt} = -\frac{M}{L_R}\frac{dI_\alpha}{dt} - \frac{M}{L_R}\omega_0 I_\beta - \frac{R_R}{L_R} i_\alpha - \omega_0 i_\beta$$

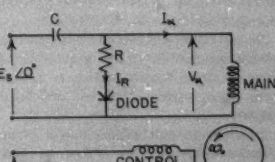


Fig. 1. Phase shift and damping circuit for a servomotor

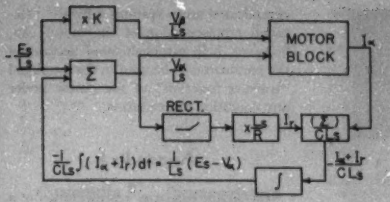


Fig. 2. Analog computer representation of nonlinear impedance in main phase

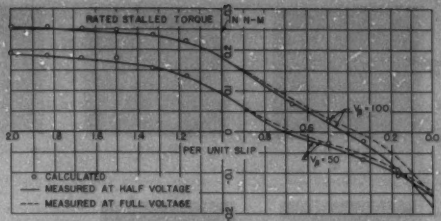


Fig. 3. Slip torque characteristics of servomotor excited through nonlinear impedance

$$\frac{di_\beta}{dt}=\frac{M}{L_R}\omega_b I_\alpha-\frac{M}{L_R}\frac{dI_\beta}{dt}+\omega_b i_\alpha-\frac{R_R}{L_R}i_\beta$$

These equations have been widely used in conjunction with operational methods to obtain transient solutions to a wide variety of problems involving constant impedances. With balanced sinusoidal excitation, these equations are greatly simplified by considerations of symmetry. The following section sets forth methods of solving these equations simultaneously with additional relationships describing the variation of source impedances with current or time.

Computer Representation of Motor and Source Impedances

An analog computer solution of this type of problem involves two parts. Part 1 relates the motor torque and current quantities to its input voltages. Part 2 of the solution solves for the impressed voltages of the motor in terms of the source emf's and impedances as well as motor currents.

Part 1, motor representation, is general. Part 2, source representation, is necessarily developed with regard to the nature of the specific system in which the motor is to be applied. In this paper, part 2 is illustrated through application to a typical diode damping circuit (see Fig. 1).

Illustrative Application

Part 1 of the computer solution is developed as shown in Appendix I. With this representation of the motor available it appears possible in the majority of cases to program the nonlinear portion of the circuit to supply the proper terminal voltages to the motor block of part 1. For the circuit of Fig. 1, this can be done in the following manner.

From an examination of the external circuits of Fig. 1, the potential drop across the phase-shift capacitor can be written as:

$$E_S-V_a=\frac{1}{C}\int_0^t(I_a+I_R)dt$$

also

$$I_R=\frac{V_a}{R}\qquad V_a>0$$

$$I_R=0\qquad V_a\leq 0$$

From Appendix I, the terminal input voltages to that portion of the computer which represents the motor are V_a/L_S and V_β/L_S.

Hence:

$$\frac{1}{L_S}(E_S-V_a)=\frac{1}{CL_S}\int_0^t(I_a+I_R)dt$$

and

$$I_R=\frac{V_a/L_S}{R/L_S}\qquad V_a>0$$

$$I_R=0\qquad V_a\leq 0$$

The program for this nonlinear external circuitry is represented in Fig. 2. The nonlinear element, the diode, is represented by a Philbrick half-wave rectifier unit. Only the I_a motor current is necessary to complete this program and this is obtained from the motor block.

The only adjustment required is the setting of the two ω_b constants to correspond to the desired speed, as shown in Appendix I. The circuit of Fig. 1 was

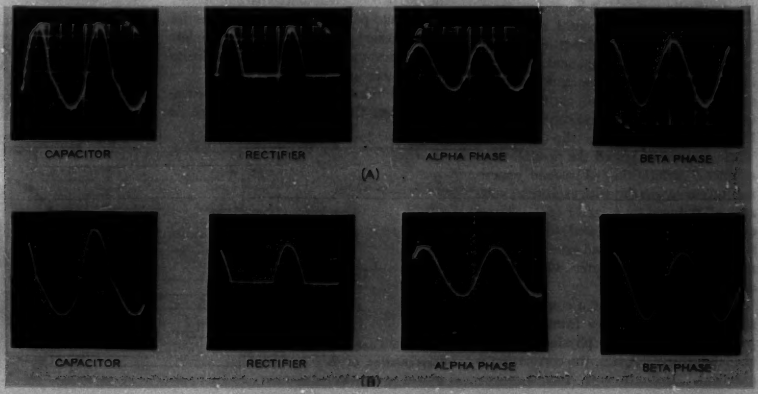


Fig. 4. Comparison of computer (A) and experimental (B) current waveforms for nonlinear driving circuit. Horizontal scales in A differ from those in B

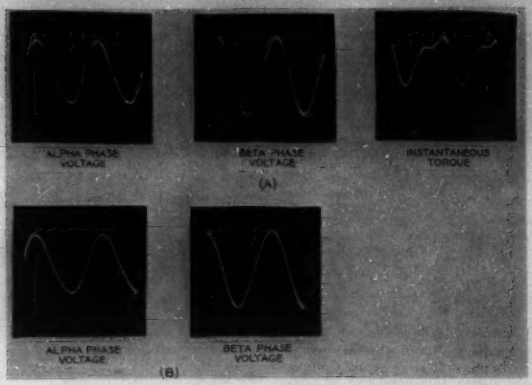


Fig. 5. Comparison of computer (A) and experimental (B) torque and voltage waveforms for nonlinear driving circuit. Horizontal scales in A differ from those in B

thoroughly investigated in the laboratory and then simulated on the Wisconsin-Philbrick analog computer.

Comparison of Experimental and Computer Results

The physical constants of the experimental circuits are presented in Appendix II, along with the scaling factors used by the authors on a fast-time computer having a 1-ms (millisecond) time base. The static slip-torque curves for the experimental and computer results are presented in Fig. 3.

The dashed curves represent the original experimental data measured with full excitation. When the computer solution was compared with these experimental results, a large discrepancy was noted in the small slip region. A careful re-examination of the experimental results revealed that both the direct component and the a-c exciting component of current in the main phase increase as slip becomes small. Also, the phase shift between currents in the main and control phases becomes very small as slip decreases. These factors suggested the possibility of magnetic saturation. A second experimental curve was run with excitation voltages all reduced by a factor of 2. These new test data were scaled to represent full voltage operation and excellent agreement with the computer solution was achieved.

Figs. 4 and 5 present a comparison of the various waveshapes from the computer solution and the physical setup in the laboratory. Again, the correlation is excellent. Instantaneous torque waveform could not be recorded in the laboratory.

Conclusions

The analysis of 2-phase servomotors excited through stator impedances which are nonlinear discontinuous functions of time and current can be accomplished through the use of the analog computer. The simple study presented herein illustrates the basic approach. More complicated studies can now be undertaken with confidence in the computer representation.

In addition, although the computer studies presented in this paper were limited to steady-state conditions, this same basic program can be readily adapted for transient studies by replacing the two ω_0 coefficient units with multipliers. This would permit a rigorous solution for the transient response of a complete carrier servomechanism, either with linear or nonlinear amplifiers.

The complexity of the motor representation makes the exact study of large or complicated servo systems difficult. Detailed representation of the motor can, however, help establish more accurately the parameters to be used in the simplified "transfer function" of the motor under any condition of excitation.

Appendix I

The analog computer representation of a 2-phase servomotor is presented in Fig. 6. This circuit is based on the physical representation of the motor, not on an equivalent circuit. The equations presented earlier will give the rotor and stator currents for the applied stator voltages, whether sinusoidal excitation is applied or not. The torque is found from the general expression for torque in terms of the space rate of change of flux linkages. Thus torque becomes:

$$T = M(I_{\beta} i_{\alpha} - I_{\alpha} i_{\beta})$$

This is instantaneous torque and will be subject to fairly violent variations if the inputs are unbalanced. The average torque may be found by putting this instantaneous torque through a lag unit having unity gain and a very long time constant.

Appendix II

The motor used for the studies undertaken in this paper is a Diehl-type *FPE 25 11*, 115-volt 60-cycle-per-second 2-phase 5-watt servomotor. This machine was found to have quite linear characteristics over its normal operating range.

The values of the motor and circuit parameters and an explanation of the various symbols used in the program are presented in Appendix III.

The actual rotor currents may be found by applying the inverse transformations from Stanley's paper. Thus, the rotor currents i_1, i_2 become:

$$i_1 = i_m \cos \omega_0 t + i_\beta \sin \omega_0 t$$

$$i_2 = -i_m \sin \omega_0 t + i_\beta \cos \omega_0 t$$

Using these physical constants in conjunction with the previously presented differential equations yields a very unwieldly set of coefficients. A set of scale factors which were found to yield a very desirable set of equations for computer representation are:

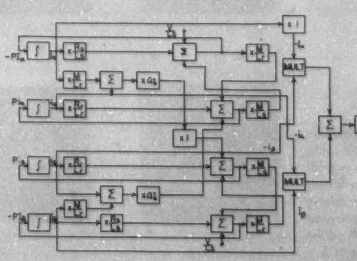


Fig. 6. Computer representation of 2-phase servomotor

$t = 0.01t'$
$I_\alpha = 0.01\ I_\alpha'$
$I_\beta = 0.01\ I_\beta'$
$I_R = 0.01\ I_R'$
$i_\alpha = 0.01\ i_\alpha'$
$i_\beta = 0.01\ i_\beta'$
$\omega_0 = 100\ \omega_0'$
$V_\alpha = V_\alpha'$
$V_\beta = V_\beta'$
$E_3 = E_3'$

Appendix III. Nomenclature

$R_R = 684$ ohms, rotor-phase resistance
$L_R = 2.84$ henrys, total rotor-phase inductance
$M = 2.59$ henrys, mutual inductance per phase
$L_S = 2.84$ henrys, total stator-phase inductance
$R_S = 317$ ohms, stator-phase resistance
$R = 333$ ohms, rectifier dropping resistor
$C = 7$ microfarads, phase-shift capacitor
ω_0 = output speed
V_α = main-phase stator voltage
$V_\beta = 100\ K$; $K = 1, 1/2$, control-phase stator voltage
$E_3 = 100$ volts, circuit input reference voltage
I_α = main-phase stator current
I_β = control-phase stator current
I_R = rectifier current
i_α, i_β = transformed rotor-phase currents

References

1. An Analysis of the Induction Machine, H. C. Stanley. *AIEE Transactions*, vol. 57, 1938, pp. 751–57.

2. Alternating Current Control of the Half-Wave Bridge Magnetic Amplifier, T. Bernstein, N. L. Schmitz. *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, Sept. 1956, pp. 433–38.

3. Operating Characteristics of 2-Phase Servomotors, E. J. W. Koopman. *Ibid.*, vol. 68, pt. I, 1949, pp. 319–29.

4. Analysis of AC Servomotors Operated from Unbalanced, Non-Sinusoidal Voltage Sources and Nonlinear Discontinuous Source Impedances, E. R. Lind. *Ph.D. Thesis*, University of Wisconsin, Madison, Wis., 1958.

Discussion

R. M. Saunders and **A. M. Hopkin** (University of California, Berkeley, Calif.): Analog computer representations for servomotors, with one exception,[1] have not received much attention in the recent technical literature although two papers on analog computer studies of the transient response of induction motors did appear about fifteen years ago.[2,3] Considering these three publications, the authors, in limiting themselves to the steady-state case, have not made a very strong contribution to the literature.

The use of direct and quadrature axis components and cross-field theory for solving servomotor problems in the time domain has the distinct advantage over the use of symmetrical components and the revolving field theory in that the equations governing the behavior of the induction machine are expressed in real rather than in complex variables. Thus the use of direct and quadrature axis components lead directly to a solution under dynamic conditions when the variables are functions of time. In general, it may be said that the price paid for obtaining the solution in real time is that simultaneous solution of the equations of motion is not as simple and straightforward as the use of the equivalent circuits which come from the symmetrical component and revolving field theory. In the authors' case it seems regrettable to pay the price of complexity without obtaining a solution valid for transient excitations.

On the other hand, the symmetrical component method yielding, as it does, equivalent circuits is admirably adapted to the solution of steady-state problems even with nonzero source impedances or even nonsinusoidal excitation. In the case of nonzero source impedances one must develop the proper sequence network quantities representing the source impedances and include them as part of the positive and negative sequence networks. For nonsinusoidal excitation one must resolve the nonsinusoidal waveform into its component harmonics by a suitable Fourier analysis and then consider the response of the network to each harmonic successively. Generally speaking, in contrast with the authors' case, the excitation of 2-phase servomotors is such that nonharmonic sources only appear in the control phase and not in the

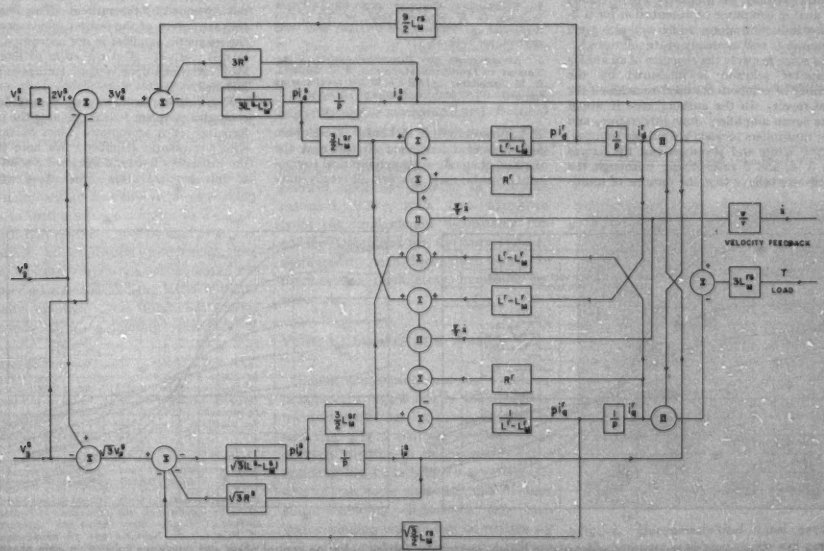


Fig. 7. Analog computer representation for a 3-phase induction motor

reference phase. Under such conditions the harmonic torque produced will have a single-phase motor characteristic (i.e., zero torque at zero speed). Further, the harmonic torques will be multiplied by $1/n^2$ at the very least. Considering the fact that the harmonic content of the voltage at the motor terminals is apt to be very low in amplitude for all components except the fundamental, only the fundamental component of the nonsinusoidal waveform needs to be considered since it is the only one which will produce an average torque. Of much more importance than the amplitudes of the harmonics is the phase shift of the fundamental component with respect to the fundamental component of the reference phase. We have found that with magnetic amplifier excitation of the control phase this effect is equally as important as the amplitude effect of the fundamental and is much more significant than the harmonic content. It is not clear in the paper if this phase shift effect is considered.

Unless saturation effects are to be considered, we can see little merit in using the direct and quadrature axis method for steady-state solutions. In the case of saturation, Slemon[4] points out that the use of direct and quadrature components on the stator and backward and forward components on the rotor yields a very satisfactory method of coping with the problem. It appears that in this particular case an analog computer representation could not be effected without the use of gyrators.

The White and Woodson[1] representation for the 2-phase servomotor is about as elemental as one can expect to achieve, and does cover not only the steady-state but the dynamic case where any time-varying excitation may be applied. Fig. 7 shows an analog computer representation for a 3-phase induction motor which will also yield dynamic as well as steady-state solutions.[5]

In some respects the elegance of an analog computer solution is measured by the number of components used to achieve the final result. In the authors' case it would take seven amplifiers, four integrators, and four multipliers to yield a dynamic solution. In the White and Woodson case the figures are 5, 4, and 6 respectively (although the discussers believe that the number of multipliers in their case could be cut to 4) and in the case of the 3-phase machine representation of Guilford and Saunders the figures are 6, 4, and 4 respectively. Even so, the use of an analog computer for 2-phase servomotors is a formidable operation and in most system studies is just not necessary.

One final note of caution: Not all servomotors are as free of space harmonics as the one shown in this paper. Erroneous results would tend to show up when the servomotor was rich in space harmonics and would tend to produce cusps in the speed-torque curve not predicted by the analog computer solution for the fundamental.

A basic weakness of the analog computer approach to problems is in the tendency of people to set up the computer before they have sufficient information concerning the true nature of the process being represented. In this case, the overlooking of saturation effects in the machine resulted in a discrepancy between computed and test data. The meaningful way to eliminate the discrepancy would have been to modify the computer representation to give a more realistic behavior of the model. The discussers note with distress that the authors eliminated the discrepancy by modifying the actual test procedure to give a less realistic measure of the motor behavior.

References

1. Electromechanical Energy Conversion (book), D. C. White, H. H. Woodson. John Wiley & Sons, Inc., New York, N. Y., 1959, pp. 486–93.

2. Transient Performance of Induction Motors, F. J. Maginniss, N. R. Schultz. *AIEE Transactions*, vol. 63, Sept. 1944, pp. 641–46.

3. Electromechanical Transient Performance of Induction Motors, C. N. Weygandt, S. Charp. *Ibid.*, vol. 65, 1946, pp. 1000–09.

4. Equivalent Circuits for Single-Phase Motors, G. R. Slemon. *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 74, 1955 (Feb. 1956 section), pp. 1335–43.

5. Applications of Electromagnetic Field Theory to Induction Machines, R. C. Guilford, R. M. Saunders. *AIEE CP 58-201* (available on request).

N. L. Schmitz and E. R. Lind: Mr. Hopkin and Mr. Saunders seem to feel that the method of steady-state symmetrical components can be used to solve all steady-state induction motor problems; even those involving nonzero source impedances which are variable discontinuous functions of current, and time. The authors do not agree with this.

It has been pointed out by the discussers that steady-state symmetrical components can be used to solve a variety of problems involving steady-state periodic voltages and constant linear impedances. This class of problem is investigated in some detail in reference 4 of the paper. Their supposition is applied conventionally to cases involving nonsinusoidal impressed emf's in one or both stator phases and in combination with d-c excitation. The fact that their supposition proved inadequate in the solution of the more general problem involving variable discontinuous source impedances led to the method reported in the paper.

The authors concur in the comments relating to harmonic torques caused by nonsinusoidal excitation through fixed stator impedances. This aspect of induction motor operation is considered in some detail in reference 4. Figs. 8(A) and (B) (Figs. 4–7 and 4–8 of reference 4) show the speed-torque characteristics of a servomotor obtained with sine-voltage excitation in the main phase and a variable length constant-amplitude voltage pulse in the control phase. In one case the pulse was started at the maximum of the sine voltage, while in the other the pulse was centered in order to keep its fundamental component in quadrature with the sine voltage. Speed-torque curves were computed using steady-state symmetrical components and correlated to test results. This technique represents a possible approach to the analysis of motor performance with magnetic amplifiers but it is not recommended by the authors. The discontinuous nature of the output impedance of the magnetic amplifier is not accounted for in this method. Also the effect of motor-load variations upon output voltage is not included.

The authors have not as yet evolved a computer program for determining the performance of a servomotor when operated with a magnetic amplifier. We note that the discussers have not disclosed a solution to this problem using either steady-state

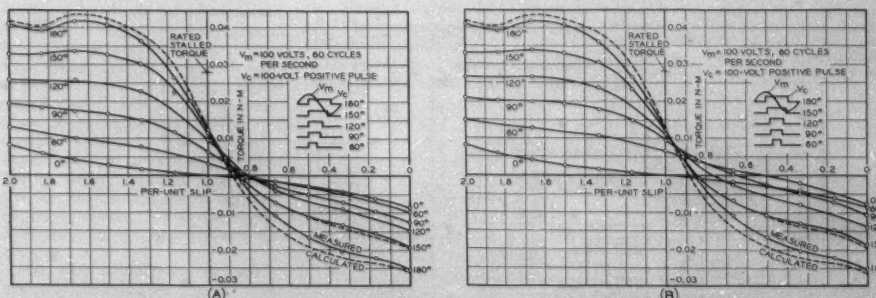


Fig. 8. Slip-torque characteristics with variable length, rectangular voltage pulses on control phase and sine-wave voltage on main phase

A—Leading edge of pulse fixed B—Fundamental component of pulse in quadrature with main-phase voltage

symmetrical components or an analog computer. This is perhaps one of the more promising areas for further investigation.

Further investigations have already indicated the feasibility of introducing multipliers, as suggested under the "Conclusions" section of the paper, for solution of transient problems. Before condemning the analog computer solution to a problem because of torque discrepancies of the order of 0.02 per unit resulting from magnetic saturation, it is perhaps advisable to inquire into the purposes of an analog computer solution. It is our belief that one of the prime merits of a fast-time computer is that of being able to observe immediately the results of changes in one or more parameters. Having established the magnitude of discrepancy introduced by certain assumptions and localized it to these assumptions, the investigator is then free to proceed with an awareness of the possible effects upon his results. The reduced voltage test reported in Fig. 3 localizes the discrepancy observed between the full-voltage test and computed results. Since the error was small (probably less than one-third that which could have resulted had not precise torque-measuring equipment and temperature-control procedures been used), it was not considered desirable to introduce variable inductance parameters to account for saturation. It is hoped that the authors' method of localizing the source of the relatively small discrepancy caused by saturation did not unduly distress our other readers.

A Suppressed-Carrier Signal Generator for Servosystem Instrumentation

P. J. POLLARD
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a method of obtaining a sinusoidal suppressed carrier voltage using two synchros connected in such a way as to eliminate the need for slip rings. The result is a relatively simple and maintenance-free signal generator that is capable of producing modulation frequencies as high as 500 cps (cycles per second).

ONE OF THE more widely accepted methods of generating a suppressed carrier signal for the analysis of servosystems is the use of "synchros." Briefly the synchro is an electromechanical device that generates a suppressed carrier signal when properly excited and rotated at some constant velocity. Being a rather rugged mechanical device, maintenance is an extremely small factor for most applications and the operation of the synchro within its limits requires nothing more than a small d-c motor as a driver. Although synchros may be used at higher carrier or excitation frequencies, the modulating frequency is a function of the speed of rotation and the latter is materially limited. This limitation corresponds to a modulating frequency of approximately 40 cps for most commercial synchros.

The advantages of using synchros, as previously pointed out, prompted the development of a method of extending the modulation frequency range. The result was the electromechanical signal generator shown in Fig. 1. The two outstanding features of this unit are:

1. It may be rotated at speeds as high as 15,000 rpm (modulating frequency is proportional to rotational speed).
2. Two cycles of the modulating frequency per revolution are obtained.

The word "cascaded" has been adopted as signifying the mechanical and electrical coupling of synchros in such a manner as to increase the initial or fundamental modulating frequency by a factor dependent upon the number of synchros cascaded. Thus with two synchros, twice the fundamental modulating frequency is obtained for the same synchro shaft speed; with three, three times the fundamental modulating frequency results; etc.

Voltage Relations in a Synchro

Before considering cascaded synchros, it is best to review briefly the equations involved with one synchro. It may be seen from Fig. 2 that the output voltage of each winding is determined by the cosine of the angle between the magnetic axis of the rotor and the axis of each stator winding. Equations for each stator winding voltage can be written by inspection, and the line-to-line voltage is the sum of the stator winding voltages between the line terminals. For an input voltage of

$$e_{in} = E \sin \omega_s t \tag{1A}$$

and making the proper assumptions as to reference and polarity, the following equations can be written for the phase voltages.

$$e_1 = kE \sin \omega_s t \cos (\theta + 180) \tag{1}$$

$$e_2 = kE \sin \omega_s t \cos (\theta + 60) \tag{2}$$

$$e_3 = kE \sin \omega_s t \cos (\theta - 60) \tag{3}$$

where $E \sin \omega_s t$ = excitation voltage on rotor, k = transformation ratio of rotor to stator, and θ = angle between rotor and stator.

The line-to-line voltages are:

$$e_{1-2} = e_1 + e_2 = kE \sin \omega_s t \cos (\theta + 120) \tag{4}$$

$$e_{2-3} = e_2 + e_3 = kE \sin \omega_s t \cos \theta \tag{5}$$

$$e_{1-3} = e_1 + e_3 = kE \sin \omega_s t \cos (\theta - 120) \tag{6}$$

The angular displacement is related to the angular velocity by the expression $\omega_m = d\theta/dt$. Thus, for a constant angular velocity ω_m, equation 5 may be rewritten as

$$e_{2-3} = kE \sin \omega_s t \cos \omega_m t$$

which can be expanded into the form

Fig. 1. Electromechanical signal generator with d-c motor as driving source and potentiometer as speed control

Paper 59-1153 recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted April 10, 1959; made available for printing August 19, 1959.

P. J. POLLARD is with Sylvania Electric Products, Inc., Detroit, Mich., and was formerly with Chrysler Corporation, Detroit, Mich.

Credit for the idea of coupling two synchros as a method of eliminating slip rings should go to D. Prichard. The author would also like to express appreciation for the technical assistance of B. Horner and L. Stopke for development of this instrument, and to Mr. Prichard for all experimental test work.

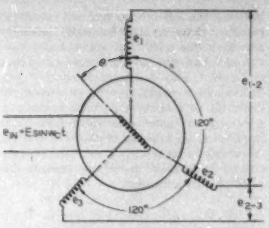


Fig. 2. Electrical schematic of synchro having a 3-phase stator and a single-phase rotor

Fig. 3. Electrical schematic of two synchros cascaded

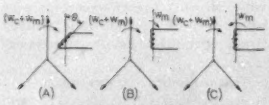


Fig. 4. Vector diagrams illustrating frequency relationships on cascaded synchros

A—Rotor fixed at angle 0, 3-phase stator excitation frequency $(\omega_c+\omega_m)$

B—Rotor angular velocity equal to ω_m and in the same direction as the 3-phase stator excitation sequence

C—Rotor angular velocity equal to ω_m and in the opposite direction to the 3-phase stator excitation sequence

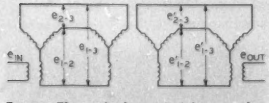


Fig. 5. Electrical schematic of three synchros cascaded

$$e_{2-3}=1/2kE[\sin(\omega_c+\omega_m)t+\sin(\omega_c-\omega_m)t] \quad (7)$$

Substitution and expansion of the other line-to-line voltages will produce the same terms with the addition of a phase angle term. From this it can be seen that in its idealized state, the synchro produces a suppressed carrier modulated signal.

Cascading Synchros

Assuming the manufacturer's rating on synchro speed is to be respected, some method of circumventing this speed limitation was needed to extend the modulating frequency range. The solution was found in the system referred to as cascaded synchros.

The electrical configuration for cascading synchros is illustrated in Fig. 3. For an input signal as expressed by equation 1(A), the three output voltages of the first synchro would be equal to those expressed by equations 4, 5, and 6, respectively. As pointed out previously, equations 4, 5, and 6 may be rewritten as

$$e_{1-3}=kE\sin\omega_c t\cos(\omega_m t+120)$$

$$e_{1-2}=kE\sin\omega_c t\cos\omega_m t$$

$$e_{1-3}=kE\sin\omega_c t\cos(\omega_m t-120)$$

Expanding these equations

$$e_{1-3}=1/2kE[\sin\{(\omega_c+\omega_m)t+120\}+\sin\{(\omega_c-\omega_m)t-120\}] \quad (8)$$

$$e_{1-2}=1/2kE[\sin(\omega_c+\omega_m)t+\sin(\omega_c-\omega_m)t] \quad (9)$$

$$e_{1-3}=1/2kE[\sin\{(\omega_c+\omega_m)t-120\}+\sin\{(\omega_c-\omega_m)t+120\}] \quad (10)$$

Upon examination of the sidebands as shown in equations 8, 9, and 10, it may be seen that both the upper and lower sidebands fulfill the time and space requirements for a 3-phase system. This fact is the basis upon which cascading synchros has been developed. When the stator of a synchro is excited with a 3-phase current the output of the rotor of the synchro will be a constant amplitude signal in which the phase relationship between the output and the input is dependent upon the rotor position. Thus, considering the 3-phase input signals of the form $\sin(\omega_c+\omega_m)t$ and $\sin(\omega_c-\omega_m)t$ to the stator of the second synchro, the output voltage would be given by the expression

$$e_{out}=\frac{3}{4}E[\cos\{(\omega_c+\omega_m)t\pm\alpha\}+\cos\{(\omega_c-\omega_m)t\mp\alpha\}] \quad (11)$$

where α is the angular position of the rotor of the second synchro with respect to its stator

At this point, the relationship of the 3-phase excitation voltage and the resultant induced voltage in the rotor winding should be studied. To simplify the explanation, only the upper 3-phase sideband will be considered. The results obtained from this will be equally applicable to the lower sideband. From the vector diagram of Fig. 4(A), the rate of change of flux linking the rotor winding and thus the frequency of the induced voltage is seen to be a function of $(\omega_c+\omega_m)$ and θ. If the rotor winding is rotated at the same angular velocity and in the same

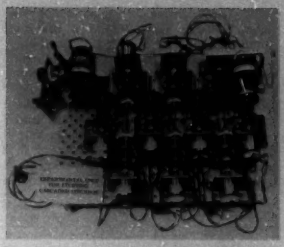


Fig. 6. Experimental model used for the study of cascading synchros

direction as the 3-phase vectors, Fig. 4(B), the resultant change in flux linking the rotor winding due to the factor ω_m becomes zero and the frequency of the resultant induced voltage will be a function of ω_c. However, if the rotor winding is rotated in the opposite direction to that of the 3-phase vectors Fig. 4(C), the rate of change of flux linking the rotor winding becomes a function of $(\omega_c+2\omega_m)$. Thus, by rotating the second synchro at the same angular velocity as the first and in the proper direction with respect to the rotation of the 3-phase vectors, the resultant output voltage will be

$$e_{out}=\frac{3}{4}E[\cos(\omega_c+2\omega_m)t+\cos(\omega_c-2\omega_m)t] \quad (12)$$

It is obvious from equation 12 that the desired doubling of the modulation frequency has been attained. It is interesting to note that the relative position of the rotor of the second synchro with respect to the rotor of the first simply determines the phase relationship of the output of the first synchro with that of the second. Thus, mechanical alignment of the respective rotors is not necessary. To cascade more than two synchros using this system, differential synchros must be used. Referring to Fig. 5 and considering an input voltage equal to $E\sin\omega_c t$, the voltages e_{1-2}, e_{2-3}, and e_{1-3} will be the same as equations 8, 9, and 10. From the previous explanation given for cascading synchros, it may be seen that the output voltages will be

$$e'_{1-2}=\frac{3}{4}kE[\cos\{(\omega_c+2\omega_m)t+120\}+\cos\{(\omega_c-2\omega_m)t-120\}] \quad (13)$$

$$e'_{2-3}=\frac{3}{4}kE[\cos(\omega_c+2\omega_m)t+\cos(\omega_c-2\omega_m)t] \quad (14)$$

$$e'_{1-3}=\frac{3}{4}kE[\cos\{(\omega_c+2\omega_m)t-120\}+\cos\{(\omega_c-2\omega_m)t+120\}] \quad (15)$$

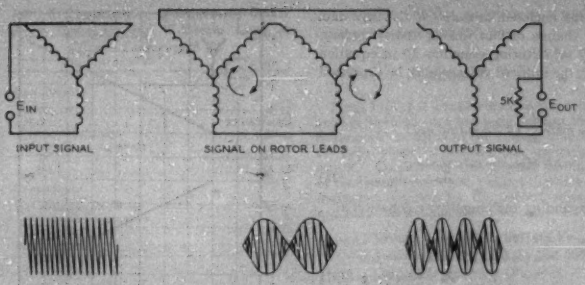


Fig. 7. Wiring diagram and corresponding waveform configurations for two synchros cascaded

This, then, becomes the 3-phase excitation for the third synchro. If the third synchro is rotated in the proper direction with respect to the 3-phase excitation, the output voltage will be

$$e_{out} = \frac{9}{8} E[\sin(\omega_c + 3\omega_m)t + \sin(\omega_c - 3\omega_m)t] \quad (16)$$

The primary limitations of the number of synchros that may be practically cascaded are the bulkiness of the resulting system, brush noise and speed regulation.

Eliminating Synchro Slip Rings

Cascading synchros had provided a method of extending the modulating frequency range but the limitations mentioned previously were considered too severe. These limitations were inflicted by the mechanical arrangement necessary; i.e., a number of synchros coupled together as shown in Fig. 6. As a result, the problem was again considered but this time with the idea of eliminating, rather than circumventing, the speed limitation of the synchro. It is evident from the previous discussion that the maximum obtainable speed of a synchro will determine its maximum modulating frequency. Since the primary factor in establishing this maximum speed rating is the slip ring and brush assembly, efforts were directed toward eliminating the latter in order that higher synchro speeds could be practically attained. Reviewing the method of cascading synchros as illustrated in Fig. 3, it may be seen that the two 3-phase stators are electrically connected. If the single-phase rotor winding and 3-phase stator winding of each synchro were interchanged; i.e., if the single-phase winding would be made the stator and the 3-phase winding would be made the rotor, operation of the synchro should not be affected. Since the rotors are mechanically coupled to a common shaft, it would then be possible to connect electrically the two 3-phase windings without the use of slip rings. The signal generator of Fig. 1 was made from two standard differential synchros (*15CDX4*) connected as shown in Fig. 7. A sleeve-type coupling was used to couple mechanically the two rotors, and lead wire was clamped along the common shaft to connect electrically the rotor windings. The resultant rotor assembly was then aligned in a cradle mount which held the two stators. Thus, the only speed limitation remaining was that of alignment and dynamic balancing. Since these could be controlled without too much difficulty a practical high-frequency suppressed-carrier signal generator was obtained.

Effect of Angular Velocity on Sideband Amplitudes

The block diagrams of Figs. 8(A) and 8(B) illustrate the instrumentation of typical Vickers electrohydraulic open-loop and closed-loop servosystems respectively. In the closed-loop system, the control signal (output of the servosystem analyzer) must be added to the feedback signal or signals. Thus, an important factor concerning the signal source when used to instrument a closed-loop servosystem is its amplitude-versus-frequency characteristic over the range to be studied. Equations 4, 5, and 6, that have been used thus far as expressing the output of a synchro, although a simplification of the true output, have been sufficiently accurate for the preceding analysis. However, a more rigorous derivation of the latter must be carried out if the amplitude versus modulating frequency characteristics of a synchro are to be understood. Thus, for an input signal given by

$$e_{in} = \sin \omega_c t \quad (1A)$$

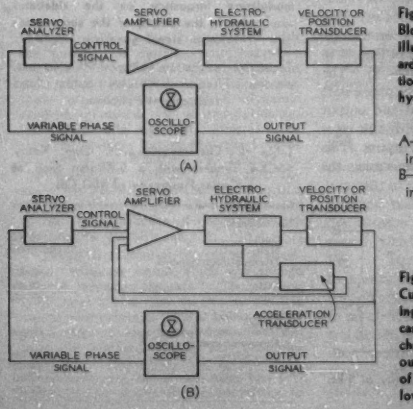


Fig. 8 (left). Block diagrams illustrating standard instrumentation for electrohydraulic servosystems

A—Open-loop instrumentation
B—Closed-loop instrumentation

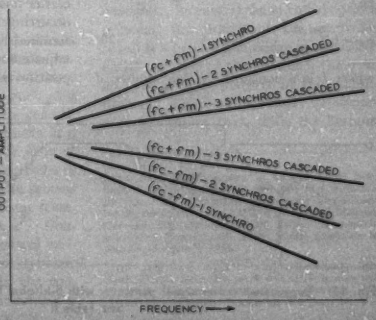

(fc + fm) – 1 SYNCHRO
(fc + fm) – 2 SYNCHROS CASCADED
(fc + fm) – 3 SYNCHROS CASCADED
(fc - fm) – 3 SYNCHROS CASCADED
(fc - fm) – 2 SYNCHROS CASCADED
(fc - fm) – 1 SYNCHRO
OUTPUT – AMPLITUDE
FREQUENCY

Fig. 9 (right). Curves illustrating the effect of cascading synchros on the output amplitude of the upper and lower sidebands

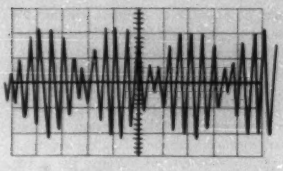


Fig. 10. Oscillogram of suppressed-carrier signal-generator output waveform frequency = 5,000 cps, modulating frequency = 500 cps, output voltage = 180 volts peak-to-peak

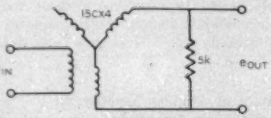


Fig. 11. Ketay 15CX4 synchro with a 5-kilohm load at a carrier frequency of 400 cps. See Table I

the instantaneous flux produced would be

$$\varphi = \Phi \cos \omega_c t \tag{17}$$

where φ is the instantaneous value of flux produced by the excitation voltage and Φ is the maximum value of flux produced by the excitation voltage.

The flux linking the secondary windings of the synchro is a function of the rotor (primary) position with respect to the stator (secondary), thus

$$\varphi' = \Phi \cos \omega_c t \sin \theta \tag{18}$$

where φ' is the instantaneous value of flux linking the secondary. Considering the synchro as rotating at some constant angular velocity equation 18 may be expressed as

$$\varphi' = \Phi \cos \omega_c t \sin \omega_m t \tag{19}$$

The voltage induced in the secondary winding under these conditions would be

$$e_{out} = kN \frac{d\varphi'}{dt} \tag{20}$$

where k is a constant, N the number of turns enclosed, and $d\varphi'/dt$ the time rate of change of flux linking the secondary. By substituting equation 19 in equation 20, the induced voltage may be expressed as

$$e_{out} = kN \frac{d}{dt} (\Phi \cos \omega_c t \sin \omega_m t)$$

$$= kN\Phi(\omega_m \cos \omega_c t \cos \omega_m t - \omega_c \sin \omega_c t \sin \omega_m t) \tag{21}$$

Expanding and simplifying equation 21

$$e_{out} = kN\Phi[1/2\ \omega_m\{\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t\} - 1/2\ \omega_c\{\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t\}]$$

$$= kN\Phi\pi[(f_c + f_m)\cos(\omega_c + \omega_m)t - (f_c - f_m)\cos(\omega_c - \omega_m)t] \tag{22}$$

It should be noted that sideband unbalance is a function of synchro speed, rather than modulating frequency. Thus in cascading synchros, the sideband unbalance characteristic changes as shown by the curves of Fig 9. If this device is to be effective in providing reliable instrumentation for closed-loop servosystems, the characteristics of the feedback elements must be considered in reference to those of the signal generator More specifically, if these characteristics differ from those of the driving signal as shown in Fig. 9, an error will result as expressed by

$$\%_{error} = \frac{f_m}{f_c}\left(1 - \frac{1}{N}\right) \times 100 \tag{23}$$

where N is the number of synchros cascaded.

Performance of Signal Generator

Figs. 10–14 indicate the performance of a signal generator. To establish a reference for comparison purposes, the characteristics of one synchro (*15Cx4*) were determined for rated excitation voltage and frequency are presented in Fig. 11 in outline form (also see Table I). The carrier phase shift refers to the phase shift occurring between the input and output terminals of the synchro with the rotor adjusted for maximum output. This position is also used to determine the

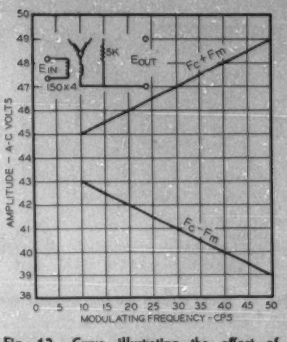


Fig. 13. Curve illustrating the effect of modulating frequency on the sideband amplitudes in the output of a Ketay 15CX4 synchro

Constants: excitation voltage = 110 volts, excitation frequency = 400 cps, output load resistance = 5 kilohms. Voltage measurements were made using the General Radio Wave Analyzer

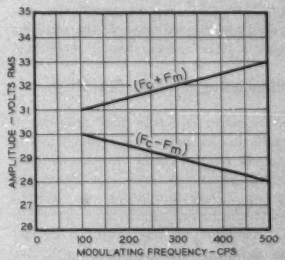


Fig. 14. Curve illustrating the effect of modulating frequency on the sideband amplitudes in the output of the suppressed-carrier signal generator

Constants: excitation voltage = 90 volts rms, excitation frequency = 5 kc, output load resistance = 5 kilohms

Table I. Characteristics of One Ketay 15CX4 Synchro with a 5-Kilohm load at a Carrier Frequency of 400 Cps

Statically:	
Gain	0.75
Null	.51 millivolts
Harmonic components	
2fc	−43 db
3fc	−49 db
4fc	−61 db
Dynamically:	
Modulating frequency	40 cps
Harmonic components	
(fc+2fm)	−36 db
(2fc+fm)	−44 db
(3fc+2fm)	−47 db
(3fc+fm)	−47 db

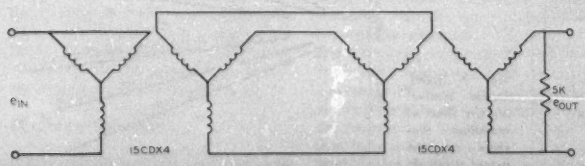


Fig. 12. Suppressed carrier signal generator with 5-kilohm load at carrier frequency of 5 kc. See Table II

Table II. Characteristics of Suppressed Carrier Signal Generator with 5-Kilohm Load at Carrier Frequency 5 Kc

Statically:	
Gain	0.71
Null	80 millivolts
Carrier phase shift	0 degrees
Harmonic components	
2/c	38 db
3/c	48 db
Dynamically:	
Modulating frequency	100 cps
Harmonic components	
2/c ± fm	35 db
3/c ± fm	48 db

harmonic components. The information of decibels (db) to facilitate comparing the operation of the subject generator to the normal operating characteristics of a synchro. Statically, the reference for the decibel ratios was the carrier frequency component present in the output with the synchro rotated to its maximum output point. Dynamically, the reference was the mean of the upper and lower fundamental sideband components at the modulating frequency indicated. The curves in Figs. 11 and 12 and Tables I and II illustrate the sideband output characteristics of a standard synchro in comparison to that of the signal generator developed. The oscillogram in Fig. 10 was taken to show the output waveform of this generator at a modulating frequency of 500 cps.

Appendix I

From Fig. 11 it can be seen that the total voltage induced into the secondary at any instant will be equal to the voltage induced by each phase of the 3-phase excitation source. Considering the angular position of the synchro rotor with respect to phase 1, a general equation for the output voltage e_{out} may be written as

$$\begin{aligned}
e_{out} &= kE[\sin\theta\sin\varphi+\sin(\theta-120)\sin(\varphi-120)+\sin(\theta+120)\sin(\varphi+120)]\\
&= 1/2\,kE[\cos(\theta-\varphi)-\cos(\theta+\varphi)+\cos(\theta-120-\varphi+120)-\cos(\theta-120+\varphi-120)+\cos(\theta+120-\varphi-120)-\cos(\theta+120+\varphi+120)]\\
&= 1/2\,kE[2\cos\theta\cos\varphi+4\sin\theta\sin\varphi-\cos\theta[\cos(\varphi+240)+\cos(\varphi-240)]+\sin\theta[\sin(\varphi+240)+\sin(\varphi-240)]]\\
&= 1/2\,kE[2\cos\theta\cos\varphi+4\sin\theta\sin\varphi+\cos\theta\cos\varphi-\sin\theta\sin\varphi]\\
&= 1/2\,kE[3\cos\theta\cos\varphi+3\sin\theta\sin\varphi]\\
&= \frac{3}{2}kE\cos(\varphi-\theta)
\end{aligned}$$

In the case of a synchro having a 3-phase carrier excitation expressed by

$$e = \sin\omega_c t$$

$$e' = \sin(\omega_c t+120)$$

$$e'' = \sin(\omega_c t-120)$$

and rotating at an angular velocity ω_m, the output voltage would be

$$e_{out} = \frac{3}{2}kE[\cos(\omega_c-\omega_m)t]$$

In the case of a synchro having a 3-phase excitation expressed by

$$e = \cos(\omega_s+\omega_m)t$$

$$e' = \cos[(\omega_s+\omega_m)t+120]$$

$$e'' = \cos[(\omega_s+\omega_m)t-120]$$

and rotating at an angular velocity ω_m, the output voltage would be

$$e_{out} = \frac{3}{2}kE[\sin(\omega_s-\omega_m)t]$$

Appendix II

From equation 22, the sideband amplitudes of a synchro may be represented as $k'(f_c+f_m)$ and $k'(f_c-f_m)$, respectively. Furthermore the sideband amplitude of the output of cascaded synchros would be $k''(f_c+f_m/N)$ and $k''(f_c-f_m/N)$ respectively, where N represents the number of synchros cascaded. If the factors k' and k'' were made equal and opposite, as in the case of negative feedback around a servoloop, the difference signal remaining would be $kf_m(1-1/N)$ and $-kf_m(1-1/N)$, respectively. Using zero modulating frequency as a reference amplitude, the per cent error introduced by the difference in the two signals would be

$$\%_{error} = \frac{f_m}{f_s}\left(1-\frac{1}{N}\right)\times 100$$

Control by Stochastic Adjustment

J. E. BERTRAM
ASSOCIATE MEMBER AIEE

Synopsis: This paper is concerned with the problem of designing an "adaptive controller" for the control of a certain class of dynamic processes. In general, these processes consist of a dynamic element governed by a nonstationary, linear differential equation. It is assumed that the designer knows only the form of the differential equation and bounds on the parameter variations but not the exact equations. A further complication of the problem is that all measurements of the state variables of the process are obscured by additive noise.

Control is exerted by means of a piecewise continuous signal which can change only at "sampling instants." The problem is to design a controller, on the basis of this meager information, which is capable of generating an input sequence that will take the process to equilibrium with zero steady-state error from any initial state. A controller capable of achieving this performance in a stochastic sense is shown to consist of a very simple multifeedback arrangement in which the only unconventional element is a time-varying gain element.

A PARTICULAR CLASS of problems of engineering and economic importance is the control, over a certain time interval, of a dynamic process such as a chemical reactor or missile. In general, the solution of this problem results in a system which has the schematic form of Fig. 1. The function of the controller is to maintain the state, $\mathbf{x}(t)$, of the process being controlled, at or near, in some sense, to an abritrary reference state, $\mathbf{x}^r$. Control action is effected by the control variable, $m(t)$, the input to the dynamic process, which is generated in the controller. Thus, the design of the controller consists of building into it some knowledge of the dynamics of the process so that, on the basis of the desired state $\mathbf{x}^r$ and the feedback information concerning the state of the process, it is possible to generate an "intelligent" control signal, $m(t)$.

In a large number of systems the dynamic behavior of the process is adequately described or approximated by an ordinary differential equation relating the state variables and the input or control variable. In most of the so-called optimal systems, the design of the controller requires a precise knowledge of the differential equation of the process if the overall system is to perform in the manner desired. This is unfortunate, for experience has shown that, while the designer

Paper 59-1156, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted February 20, 1959; made available for printing August 6, 1959.

J. E. BERTRAM, with Columbia University, New York, N. Y., when this work was done, is now with Research Center, International Business Machines Corporation, Yorktown Heights, N. Y.

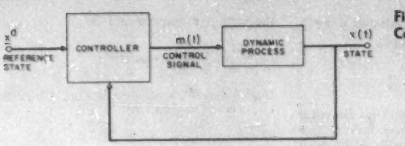


Fig. 1. (left). Control system

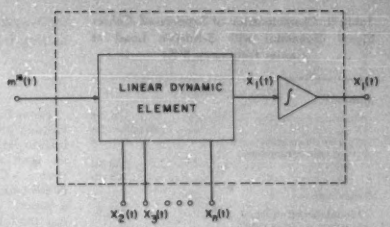


Fig. 2. (right). Dynamic process

may know the order and form of the equation, he seldom knows with any precision the exact values of the parameters. Even when measurements are made and precise estimates of the parameters are obtained prior to designing the controller, it has been found in most applications that these parameters change appreciably during the life of the system either because of aging of certain critical components or because of environmental changes.

In the past few years a considerable effort has been made to develop control-system design methods which do not require that the designer have a precise a priori knowledge of the dynamics of the process.

These efforts, in general, have been based more on intuition than on analysis. Usually the procedure has been to select some criterion of performance which is readily measurable, perform this measurement continuously on the system, and utilize the measurement to adjust certain free parameters in the controller in order to achieve the desired performance. Examples of such systems are found in the work of Draper and Li, Burt,[1] Cosgriff,[2] and also Anderson and Aseltine.[3] Kalman[4] has suggested a somewhat different approach to the problem which is applicable to the control of any dynamic process which is adequately described by an ordinary, linear, differential equation which changes slowly with time. In his system, the process dynamics are estimated from input–output measurements and on the basis of these estimates the controller parameters are calculated and adjusted in compliance with some performance criterion. In other words, the standard design procedure is mechanized. At various times these systems have been classified as "self-adjusting," "self-optimalizing," or "adaptive." In the following these terms are considered as synonymous.

In each of the systems mentioned before it is assumed that any time variation in the process parameters is slow in comparison with either the response time of the process or the measurements and adjustment time of the controller. Furthermore, since the dynamic performance of these systems with the self-adjustment loops is described by nonlinear differential or difference equations, little is known about the transient behavior. In those systems which have been built[3,4] it has been found that satisfactory performance has required the introduction of a disturbance or test signal in order to insure a minimum signal-to-noise ratio in the measurements. This signal limits the steady-state accuracy of the system.

This paper is concerned with a different approach to the adaptive control problem. The methods employed are strongly motivated by and are closely related to an area of current interest in mathematical statistics called stochastic approximations. In their pioneering paper, Robbins and Monro[5] give an algorithm for stochastically converging on the solution, x, of the equation $M(x)=\alpha$, where $M(x)$ is the expected value of the response of an experiment to the input x. (It is assumed that the function $M(x)$ is unknown to the experimenter.) Thus, the problem is to find that input x which on the average yields the output α. Their results have been extended by numerous authors [6–8] A point of view which is of particular interest in adaptive control systems was suggested by Dvoretzky.[9] He considered the stochastic approximation method as a convergent deterministic procedure with a superimposed random element or noise. The adaptive-control problem of this paper is viewed in a similar way.

Statement of the Problem

The dynamic process to be controlled is illustrated in Fig. 2. The following assumptions concerning the process are made:

1. It is assumed that the input, $m^*(t)$, generated in the controller, is sampled-and-held so that it is piecewise constant. That is, between successive sampling instants the signal is held constant at the value of the function at the previous sampling instant. Thus:

$$m^*(t)=m^*(kT),\ kT\leq t<(k+1)T,\quad k=0,\ 1,\ 2,\ \ldots \qquad (1)$$

where the sampling period, T, is a positive constant and the point, $t=kT$, is called the kth sampling instant. (Frequently it will be convenient to write $f(k)$ rather than $f[kT]$.)

2. The state variables of the dynamic process denoted by $x_1,\ x_2,\ x_3,\ \ldots,\ x_n$ are observed or measured only at the same sampling instants ($t=0,\ T,\ 2T,\ \ldots$). The state of a dynamic element at any instant of time is a set of numbers which completely summarize that past history of a dynamic element which is necessary to a determination of its future behavior. To be specific, if one thinks of the element as being simulated on an analog computer, then the voltages at the outputs of the integrators form such a set. (The number of integrators needed to simulate an nth order differential equation is exactly n and no more.) Since the simulation of a dynamic element is not unique, it follows that the state variables are not unique. However, in the present case it is assumed that the state variables have been selected so that all n of them are physically accessible to measurement. A discussion of how this condition may be relaxed is found in reference 10.

3. It is assumed that in the process of performing the observation or measurement a random element (noise) is added to the state variable so that:

$$y_i(t)=x_i(t)+n_i(t),\ i=1,\ 2,\ \ldots,\ n \qquad (2)$$

where y is the observation, x is the true state value, and n is the noise.

4. The dynamic process contains an integrator so arranged that if the differential equations describing the process are written in normal form,[11] then the derivatives of the state variables with respect to time are functions only of the state variables $x_2,\ x_3,\ \ldots,\ x_n$, the input m^*, and time, so that:

$$\dot{x}_i(t)=\sum_{j=2}^{n} a_{ij}(t)x_j(t)+d_i(t)m^*(t),\quad i=1,\ 2,\ \ldots,\ n \qquad (3)$$

Consequently, equation 3 is such that the only equilibrium states for $m^*=0$ are:

$$x_1=x_1^r \text{ (an arbitrary constant)}$$
$$x_i=0,\ i=2,\ 3,\ \ldots,\ n \qquad (4)$$

The restrictions to sampling which are inherent in assumptions 1 and 2 only tend to simplify the problem without any real loss in generality. Assumption 3 merely states a condition which is always present in the physical world. On the other

hand, 4 somewhat limits the applicability of the result, but fortunately this restricted class includes a large number of physical problems. Actually this assumption may be relaxed considerably; in the present paper this has not been done so that the mathematical arguments may be kept simple in order not to obscure the motivation and methods employed.

The problem now is to determine under what conditions, if any, it is possible to design a controller which will generate a control sequence $m(0)$, $m(1)$, ... which will take the process from any initial state $\mathbf{x}(0)$ to a desired equilibrium state $\mathbf{x}^r$ in some optimal fashion. Further, it is desired to use a simple controller and one which can be designed from as little a priori knowledge concerning the time-varying parameters and the noise, n_i, as possible.

The state of the art at the present time is not at the point where a unique best solution is possible. This paper presents a solution which satisfies most of the conditions of the problem and has the potential of satisfying many more.

Description of Over-All System

If the state variables are observed only at the sampling instants, $t=0, T, 2T, \ldots$, then the solution of equation 3 will yield the difference equation:

$$\mathbf{x}(k+1)=\Phi_k\mathbf{x}(k)+\mathbf{d}_k m^*(k) \tag{5}$$

where $\mathbf{x}(k)$ is an n-dimensional vector with components $(x_1, x_2, \ldots, x_n)$ denoting the state of the process at the kth sampling instant; Φ_k is an $(n\times n)$ square matrix which represents the transformation of the state at the kth to $(k+1)$th sampling instant in the absence of an input; $\mathbf{d}_k$ is an n vector which represents the effect on the state at the $(k+1)$th instant of a unit step-control signal applied at the kth instant; and $m^*(k)$ is a scaler denoting the value of the kth control signal.

The adaptive controller suggested in this paper is a simple device which generates the control signal $m^*(k)$ by weighting the difference between the desired equilibrium state $\mathbf{x}^r$ and the observed state vector $\mathbf{y}(k)$ by the time-varying gain $\mathbf{a}_k$. Thus:

$$m^*(k)=\mathbf{a}_k'[\mathbf{x}^r-\mathbf{y}(k)] \tag{6}$$

where $\mathbf{a}_k'$ is the transpose of vector $\mathbf{a}_k$. The observed state vector $\mathbf{y}(k)$ is actually the sum of a noise vector $\mathbf{n}(k)$ and the actual state vector $\mathbf{x}(k)$.

$$\mathbf{y}(k)=\mathbf{x}(k)+\mathbf{n}(k) \tag{7}$$

Equations 5, 6, and 7 describe the over-all system shown in Fig. 3. To obtain

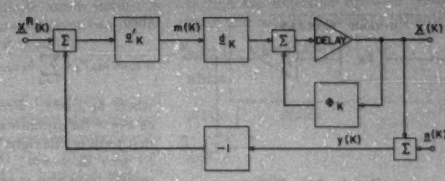


Fig. 3. Linear system

the over-all difference equation describing the behavior of this system at sampling instants, substitute equations 6 and 7 into 5 and find:

$$\mathbf{x}(k+1)=(\Phi_k-\mathbf{d}_k\mathbf{a}_k')\mathbf{x}(k)+\mathbf{d}_k\mathbf{a}_k'\mathbf{x}^r-\mathbf{d}_k\mathbf{a}_k'\mathbf{n}(k) \tag{8}$$

The problem now is to show that under conditions which are reasonable in practice, $\mathbf{x}(k)$ in some sense tends to the desired equilibrium state, $\mathbf{x}^r$. Rather than investigate equation 8 for all $\mathbf{x}^r$, it is easier to subtract $\mathbf{x}_k{}^r$ from both sides of 8, and letting $\mathbf{e}(k)=\mathbf{x}(k)-\mathbf{x}^r$, find conditions under which $\mathbf{e}(k)$ tends to zero as k becomes large. This result can be written:

$$\mathbf{e}(k+1)=(\Phi_k-\mathbf{d}_k\mathbf{a}_k')\mathbf{e}(k)-(\Phi_k-I)\mathbf{x}^r-\mathbf{d}_k\mathbf{a}_k'\mathbf{n}(k) \tag{9}$$

where I denotes the identity matrix (unity on the diagonal, zero elsewhere). However, from condition 4, which is explicitly stated in equations 3 and 4, the term

$$(\Phi_k-I)\mathbf{x}^r=0 \tag{10}$$

From equation 3 the first column of the matrix Φ_k contains only zeros except for the Φ_{11} term which is always unity. Consequently the matrix (Φ_k-I) has only zeros in the first column. From equation 4 the equilibrium reference vector $\mathbf{x}^r$ has all components equal to zero except $x_1{}^r$ which is arbitrary. That is:

$$\mathbf{x}^r=\begin{bmatrix} x_1{}^r \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{11}$$

As a result, the product of equation 10 is identically equal to zero and equation 9 is simply:

$$\mathbf{e}(k+1)=(\Phi_k-\mathbf{d}_k\mathbf{a}_k')\mathbf{e}(k)-\mathbf{d}_k\mathbf{a}_k'\mathbf{n}(k) \tag{12}$$

Conditions for Convergence

The conditions for convergence of equation 12 may be stated formally as follows:

In the equation the expected or mean value of the norm of $\mathbf{e}(k)$ denoted by $E\|\mathbf{e}(k)\|$ tends to zero as $k\to\infty$

$$\lim_{k\to\infty} E(\|\mathbf{x}(k)-\mathbf{x}^r\|)\to 0 \tag{13}$$

if the following conditions are satisfied:

1. The norm

$$\|\Phi_k-\mathbf{d}_k\mathbf{a}_k'\|<\mu_k \tag{13A}$$

where μ_k is a sequence of positive numbers satisfying

$$\prod_{k=0}^{\infty}\mu_k=0 \tag{13B}$$

and such that the partial products of equation 13(B) are uniformly bounded.

2. The $\mathbf{n}(k)$ for $k=0, 1, 2, \ldots$ are random vectors which are statistically independent of the dynamic system and such that the expected value of their norm is finite:

$$E(\|\mathbf{n}(k)\|)<\infty \tag{13C}$$

3. The infinite sum:

$$\sum_{k=0}^{\infty}E(\|\mathbf{d}_k\mathbf{a}_k'\|)E(\|\mathbf{n}(k)\|) \tag{13D}$$

is finite.

The proof and application of this result require at least an elementary understanding of the concept of norm. The norm of a vector $\mathbf{y}$ written as $\|\mathbf{y}\|$ is a scaler quantity denoting in some sense the magnitude or length of $\mathbf{y}$. Examples of a norm are:

$$\|\mathbf{y}\|_1=\sum_{i=1}^{n}|y_i| \tag{14A}$$

$$\|\mathbf{y}\|_2=\left[\sum_{i=1}^{n}(y_i{}^2)\right]^{1/2} \tag{14B}$$

or

$$\|\mathbf{y}\|_3=\max_i(\alpha_i|y_i|) \tag{14C}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are a set of positive numbers.

It is readily verified that each of these norms has the properties that:

$$\|\mathbf{x}+\mathbf{y}\|\leq\|\mathbf{x}\|+\|\mathbf{y}\| \tag{15A}$$

$$\|c\mathbf{y}\|=|c|\ \|\mathbf{y}\| \tag{15B}$$

$\|\mathbf{y}\|>0$ if $\mathbf{y}\neq 0$

$\|\mathbf{y}\|=0$ if, and only if, $\mathbf{y}=0$ (15C)

($\mathbf{y}=0$ implies that every component of $\mathbf{y}$ is zero)

In defining matrix norms it is natural to impose the same conditions and possibly a few others;[11] that is:

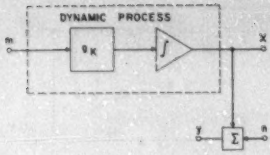


Fig. 4. Example 1

$$\|A+B\| \leq \|A\| + \|B\| \tag{16A}$$

$$\|\alpha A\| = |\alpha| \, \|A\| \tag{16B}$$

where α is a scaler, and

$$\|A\| \neq 0 \text{ unless } A = 0 \tag{16C}$$

A particular matrix norm will be said to be consistent with a given vector norm if for every A and $\mathbf{x}$:

$$\|A\mathbf{x}\| \leq \|A\| \, \|\mathbf{x}\| \tag{16D}$$

Matrix norms which satisfy equations 16(A), 16(B), and 16(C) and which are consistent respectively with the vector norms of equation 14 are:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}| \tag{17A}$$

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij})^2 \right]^{1/2} \tag{17B}$$

Either

$$\max_i \left(\sum_{j=1}^{n} \frac{\alpha_i}{\alpha_j} |a_{ij}| \right) \tag{17C}$$

or

$$\max_j \left(\sum_{j=1}^{n} \frac{\alpha_j}{\alpha_i} |a_{ij}| \right)$$

The proof of equation 13 is quite straightforward using the concept of norm and will be presented here because it illustrates the applicability and limitations of the result.

The first step is to take the norm of equation 12 and use equations 15(A) and 16(D) to obtain:

$$\|\mathbf{e}(k+1)\| < \|\phi_k - \mathbf{d}_k \mathbf{a}_k'\| \, \|\mathbf{e}(k)\| + \|\mathbf{d}_k \mathbf{a}_k'\| \, \|\mathbf{n}(k)\| \tag{18}$$

Next take the ensemble average or expected value of equation 18. Letting $V_k = E(\|\mathbf{e}(k)\|)$ and $\sigma_k = E(\|\mathbf{d}_k \mathbf{a}_k'\|) \, E(\|\mathbf{n}(k)\|)$ and using condition 13(A):

$$V_{k+1} < \mu_k V_k + \sigma_k \tag{19}$$

is found, a first-order difference equation. Upon successively iterating equation 19, let

$$b_{km} = \prod_{j=m}^{k} \mu_j = \mu_m \mu_{m+1} \ldots \mu_k$$

then:

$$V_{k+1} < b_{k0} V_0 + \sum_{i=0}^{k} b_{k,i+1} \sigma_i \tag{20}$$

where $b_{k,k+1} = 1$. From equation 13(B) as $k \to \infty$ for any finite m, $b_{km} = 0$ so that the first term on the right in equation 20 goes to zero as $k \to \infty$. To show that the second term also goes to zero as $k \to \infty$, break it into two partial sums:

$$\sum_{i=0}^{k} b_{k,i+1} \sigma_i = \sum_{i=0}^{m} b_{k,i+1} \sigma_i + \sum_{i=m+1}^{k} b_{k,i+1} \sigma_i \tag{21}$$

for any finite m, it follows from equation 13(B) that as $k \to \infty$, each term in the first sum of equation 21 goes to zero. Further, from 13(B) all partial products of b_{ij} are uniformly bounded by some finite positive constant. Since from equations 13(C) and 13(D)

$$\sum_{k=0}^{\infty} \sigma_k < \infty$$

it is possible to choose an m in equation 21 so that:

$$\sum_{i=m+1}^{\infty} b_{k,i+1} \sigma_i < \epsilon$$

for any positive ϵ. Therefore, it has been shown that in the limit as $k \to \infty$, V_k converges to zero.

Note that in control systems one is generally interested in designing systems which are asymptotically-stable-in-the-large (i.e., for any initial state $\mathbf{x}(0)$ the norm of the solution $\mathbf{x}(k)$ tends to the equilibrium point as $k \to \infty$). In this case however, the best solution achieved is that the expected value of the norm, $E(\|\mathbf{x}(k) - \mathbf{x}^r\|) \to 0$ as $k \to \infty$. Roughly speaking, this means that $\mathbf{x}(k)$ generally converges to $\mathbf{x}^r$ but it is not an absolute certainty. While this limitation is not desirable, it seems to be the best that can be achieved for the generality of the present case.

To illustrate how this result is applied to the design of a particular system, consider the following first-order example.

Example 1

For the dynamic process of Fig. 4, it is desired to determine an input sequence $m(0)$, $m(1)$, ... which will bring an arbitrary initial state $x(0)$ to the desired state x^r. The process consists of an integrator preceded by a time-varying gain element $g(t)$. The only a priori information available to the designer is that these variations are bounded by the positive constants α_1 and α_2 such that $\alpha_1 < g(t) < \alpha_2$.

The differential equation describing the process is:

$$\frac{dx}{dt} = g(t) m(t) \tag{22}$$

Since the input is a piecewise constant signal (defined by equation 1), the difference equation relating the state variable, x, and the input m at sampling instants is:

$$x(k+1) = x(k) + d_k m(k) \tag{23}$$

where

$$d_k = \int_{kT}^{(k+1)T} g(\tau) d\tau$$

Consequently, $\alpha_1 T < d_k < \alpha_2 T$.

The observation of the state variable $y(k)$ is a linear combination of the actual state variable $x(k)$ and a random element or noise, $n(k)$. The only information available about this noise is that its expected value or mean is zero, while the variance is bounded by a positive constant. That is $E[n(k)] = 0$ while $E(n(k)^2) < \infty$ for all k.

From this information it is possible to generate an input sequence $m(0)$, $m(1)$, ..., which insures convergence in the sense described by simply multiplying the difference between the desired state x^r and the observation $y(k)$ by the time-varying gain $a_k = A/(1+k)$ where $A > 0$, and using the result as a control signal. A wise choice of A speeds the convergence.

To see that this simple strategy works, write the difference equation of the complete system of Fig. 5.

$$(x(k+1) - x^r) = \left(1 - \frac{A}{1+k} d_k\right)(x(k) - x^r) - \frac{A}{1+k} d_k n(k) \tag{24}$$

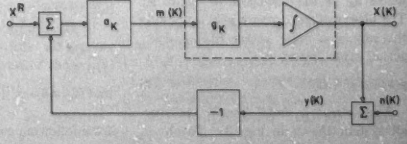


Fig. 5. The system

In the first-order system the state vector has only one component. If it is desired to obtain convergence in the mean square, x^2 can be used rather than the norm x for this example. If the system is to converge in the mean square:

$$\lim_{k\to\infty} E(x(k)-x^r)^2 \to 0$$

then it follows from equations 13(A)–(D) that, first:

$$\left(1-\frac{A}{1+k}\,d_k\right)^2 < \mu_k \tag{25A}$$

where μ_k is a sequence of positive numbers such that

$$\prod_{k=0}^{\infty} \mu_k = 0 \tag{25B}$$

A necessary and sufficient condition[13] that the infinite product:

$$\prod_{k=0}^{\infty}\left(1-\frac{A}{1+k}\,d_k\right)=0$$

is that the infinite sum

$$\sum_{k=0}^{\infty}\frac{A}{1+k}\,d_k = \infty$$

Obviously this condition is satisfied since

$$Ad_k < c_3\, TA$$

and

$$\sum_{k=0}^{\infty}\frac{1}{1+k} = \infty$$

Second, it is also necessary that

$$\sum_{k=0}^{\infty} E\left(\left(d_k\,\frac{A}{1+k}\right)^2\right) E(n(k)^2) < \infty \tag{25C}$$

be finite. By assumption $E\ (n(k)^2)$ is finite and $d_k A < c_3 TA$ which is finite. It follows that equation 25(C) is less than

$$C\sum_{k=0}^{\infty}\frac{1}{(1+k)^2} < \infty \tag{26}$$

where C is a finite positive constant, since the infinite sum of equation 26 is kown to be finite.

Thus in this system with the gain, $g(t)$, varying in an unknown but bounded fashion, and with the feedback measurements obscured by noise, it is possible to generate a control or input sequence $m(0)$, $m(1)$, ..., which takes an initial stage, $x(0)$, to the desired state x^r in the stochastic sense (i.e., $E(x(k)-x^r)^2 \to 0$ as $k\to\infty$) without the steady-state variance of error usually associated with noise in a system.

Note that to satisfy equation 25(A) and (B) the time-varying gain cannot tend to zero too rapidly, but in order to satisfy 25(C) its square must tend to zero fast enough that its infinite sum is finite. A sequence which satisfies these two conditions is classified as type $1/k$.

Comparison

As a comparison of this design method with that of conventional sampling systems, consider the previous example with $T=1$ second, $g(t)=10$; $E(n(kT)^2)=\sigma^2$, and the noise, $n(kT)$, "sampled white."[18]

If this system is designed in the conventional way for "finite settling time" in the absence of noise, the controller is a constant gain, $a(k)=0.1$. In the presence of noise, the variance of the system error for all k is $E(x(k)-x^r)^2=\sigma^2$.

In comparison, if the method of this paper is used and the controller gain $a(k)$ is made equal to $0.1/(1+k)$, then the variance of system error as a function of k is:

$$E(x(k)-x^r)^2 = \frac{\sigma^2}{1+k}$$

Thus for all $k>1$ the variance of the latter system error is less than one-half that of the conventional sampling system and, further, it tends to zero with large k.

Practical Considerations

First, since $a(k)\to 0$ as $k\to\infty$, it would seem that such a system would only be useful as a one-shot device. To get around this difficulty and permit continuous performance, it is necessary to add an element to the controller which starts the sequence $a(k)$ over again when a change in the reference state occurs. If the reference state, x^r, changes infrequently, that is, at a rate which is much slower than the relative convergence rate of the system, then the performance is the same as that previously described. If the input changes are rapid, then the present system is not adequate for the task.

A second and more serious problem concerns the effect of a disturbance at the input of the integrator. With such a disturbance it is not possible to achieve convergence in the sense of

$$\lim_{k\to\infty} E(x(k)-x^r)^2 \to 0$$

nor could such convergence be obtained by any other scheme as long as the control effort is restricted to manipulating the input, $m(t)$.

To illustrate the effect of such a disturbance assume that in the previous example a disturbance $q(t)$ is added to the input of the integrator The equation governing the dynamic element now takes the form:

$$\dot{x} = d(t)m^*(t) + q(t)$$

so that the difference equation relating input and output at sampling instants is:

$$x(k+1) = x(k) + d(k)m^*(k) + q'(k)$$

where

$$d(k) = \int_{kT}^{(k+1)T} d(\mu)\,d\mu$$

and

$$q'(k) = \int_{kT}^{(k+1)T} q(\mu)\,d\mu$$

Using the controller of the previous section, the system equation now takes the form

$$e(k+1) = (1-a(k)d(k))e(k) - a(k)d(k)m(k) + q'(k) \tag{27}$$

This equation without the disturbance is equation 24.

From equation 27 it can be seen that for a constant disturbance $q(t)=Q$, $(q'(k)=QT)$ the system is such that the deviation $e=x-x^r$ grows without bound. The expression for this deviation due to the disturbance alone $(n(k)=e(0)=0)$ is:

$$e(k+1) = \sum_{j=0}^{k} f_{k,j+1}q'(j)$$

where $f_{k,j+1} = [1-a(j+1)d(j+1)]\ldots[(1-a(k)d(k)]$. For $a(k)$ of the form $A/(1+k)$ the infinite sum:

$$\lim_{k\to\infty}\sum_{j=0}^{k} f_{k,j+1} = \infty$$

does not exist.

If the disturbance, however, is random, independent of the observation noise and the state of the dynamic process, has mean $E(q'(k))=0$, and has a variance $E(q'(k)^2)$ which is finite; then the effect is quite different. Letting $U_k = E(e(k)^2)$, $\sigma_k = E[a(k)d(k)(n(k))]$, and $Q(k) = E(q'(k)^2)$ the following is obtained:

$$V_{k+1} < \mu_k V_k + \sigma_k + Q_k$$

where $\mu_k < (1-a(k)d(k))^2$. Successively iterating from $k=0$, find:

$$V_{k+1} < b_{k,0}\,V_0 + \sum_{j=0}^{k} b_{k,j+1}\sigma_j + \sum_{j=0}^{k} b_{k,j+1}Q_{(j)}$$

where $b_{k,j}$, as previously defined, is the product $\mu_j\mu_{j+1}\cdots\mu_k$.

From previous results it is known that that the first two terms on the right of equation 28 tend to zero as k becomes large. The third term:

$$\sum_{j=0}^{k} b_{k,\,j+1}Q_{(j)} < C\sum_{j=0}^{k} b_{k,\,j+1} < \infty$$

Table I. Transient Response

			x(K)				
K	g_K	n_K	A=1	=1/4	=1/8	=1/16	=1/32
00000	+000.000	+000.000	+020.000	+020.000	+020.000	+020.000	+020.000
00001	+007.978	+005.600	+020.000	+020.000	+020.000	+020.000	+020.000
00002	+006.162	+006.920	−184.256	−031.064	−005.532	+008.185	+007.233
00003	+009.569	−006.600	+116.834	−012.465	−006.066	+005.908	+004.508
00004	+009.984	−004.051	−234.792	+002.738	−001.016	+006.266	+004.925
00005	+011.038	+004.957	+100.319	+003.558	+000.565	+006.123	+004.788
00006	+007.541	−001.986	−132.090	−001.141	−000.958	+004.520	+003.444
00007	+008.460	−001.419	+036.438	−000.158	−000.495	+004.662	+003.320
00008	+002.193	−000.822	−005.888	+000.318	−000.206	+004.525	+003.185
00009	+007.079	+001.004	−004.048	+000.352	−000.171	+004.543	+003.144
00010	+003.025	+004.612	−001.654	+000.085	−000.253	+004.240	+002.940
00011	+014.123	−001.698	−002.549	−000.269	−000.418	+004.067	+002.798
00012	+003.027	+003.099	+002.905	+000.362	−000.078	+003.765	+002.709
00013	+013.308	+002.469	+001.390	+000.143	−000.173	+003.685	+002.618
00014	+011.929	+007.207	−002.561	−000.524	−000.467	+003.386	+002.292
00015	+013.674	−001.439	−006.520	−001.948	−001.185	+003.014	+001.786
00016	+007.618	+001.268	+000.736	−001.176	−000.886	+003.182	+001.766
00017	+008.093	−006.611	−000.218	−001.187	−000.908	+002.963	+001.676
00018	+013.308	+006.405	+003.033	−000.259	−000.461	+003.098	+001.823
00019	+014.335	−003.320	−003.945	−001.395	−001.010	+002.715	+001.443
00020	+006.795	+003.327	+001.537	−000.505	−000.602	+002.720	+001.531
00021	+012.860	−000.055	−000.115	−000.745	−000.717	+002.552	+001.428
00022	+015.716	−006.001	−000.010	−000.622	−000.658	+002.524	+001.376
00023	+013.888	+006.038	+004.284	+000.560	−000.064	+002.482	+001.582
00024	+010.177	+003.307	−001.949	−000.435	−000.514	+002.287	+001.294
00025	+010.352	−004.780	−002.525	−000.740	−000.662	+002.153	+001.172
00026	+014.999	−005.483	+000.500	−000.168	−000.381	+002.111	+001.266
00027	+000.628	+001.337	−003.379	+000.646	+000.041	+002.066	+001.418
00028	+010.096	+006.403	+003.265	+000.634	+000.037	+002.053	+000.414
00029	+010.699	+001.017	−000.221	+000.000	−000.252	+001.973	+001.238
00030	+000.607	−005.761	−000.514	−000.093	−000.287	+001.812	+001.186
00031	+008.674	−005.834	−000.387	−000.063	−000.272	+001.818	+001.191
00032	+006.511	−003.290	+001.353	+000.348	−000.058	+001.869	+001.273
00033	+014.695	+003.883	+001.747	+000.498	+000.026	+001.935	+001.298
00034	+010.031	+004.248	−000.759	+000.010	−000.191	+001.810	+001.154
00035	+012.777	−005.281	−001.789	−000.303	−000.340	+001.708	+001.054
00036	+008.128	+004.080	+000.792	+000.206	−000.084	+001.755	+001.151
00037	+003.438	+004.985	−000.307	−000.035	−000.197	+001.709	+001.077
00038	+002.497	+002.969	−000.742	−000.150	−000.252	+001.673	+001.042
00039	+008.466	+001.857	−000.888	−000.198	−000.275	+001.645	+001.025
00040	−000.491	−000.172	−001.099	−000.287	−000.318	+001.613	+000.986
00041	+003.201	+000.023	−001.083	−000.285	−000.317	+001.612	+000.986
00042	+006.401	+001.572	−001.000	−000.280	−000.314	+001.587	+000.981
00043	+006.114	−001.749	−001.087	−000.329	−000.338	+001.520	+000.956
00044	+004.179	+001.869	−000.684	−000.255	−000.301	+001.540	+000.963
00045	+006.860	−003.499	−000.797	−000.294	−000.310	+001.490	+000.947
00046	+006.121	+000.347	−000.141	−000.149	−000.247	+001.513	+000.971
00047	+013.767	−003.001	−000.169	−000.156	−000.248	+001.479	+000.960
00048	+001.020	−000.839	+000.759	+000.075	−000.129	+001.593	+000.997
00049	+008.793	−002.077	+000.761	+000.079	−000.127	+001.599	+000.997
00050	+000.686	+000.559	+000.997	+000.168	−000.077	+001.611	+001.009

In Table I, the transient performance of the system described, with an initial state $x(0)=20$, sampling period of $T=1$ second, and the noise and gain sequences tabulated, is shown. The terms in the gain sequence are confined to the range, $0.1<g(k)<16$, while the noise $n(k)$, is uniformly distributed in the range from -8 to $+8$ (i.e., $E[n(k)]=0$, $E[n(K)]^2=4/3$). The response is tabulated for $A=1$, $1/8$, $1/16$, and $1/32$. The latter two choices insure that μ_k is always less than unity, but the fastest convergence is obtained with $A=1/8$.

Example 2

Example 2 is similar to example 1 except that it is of second order rather than first order (see Fig. 6).

The information available to the designer is the following:

1. The time-varying gain $\alpha<g_K<\beta$ where α and β are positive numbers.
2. The noise vector $\mathbf{n}(k)$ is such that $E[\|\mathbf{n}(k)\|]<\infty$ for all k.
3. The parameter, a, is positive.

Using the vector norm 15(C) and its consistent matrix norm 17(C), it can be shown that $a_{ik}=A_i/(1+k)^2$, where A_i is a positive constant, satisfies equations 13(A)–(D), and thus insures that:

$$E(\|\mathbf{x}(k)-\mathbf{x}^T\|)\to 0$$

in the limit as $k\to\infty$. As before, the transient behavior is closely related to the choice of A_i.

Conclusions

This paper presents a method for designing a control system when the process being controlled has the following properties:

1. Parameters varying in a bounded fashion.
2. The measurements of the state variables are obscured by additive noise.

The controller required to insure stability and a transient response which is relatively insensitive to the parameter variations and noise, is shown to be a nonlinear, time-varying, digital filter. The filter is described by a set of nonlinear, first-order, difference equations. The coefficients in these equations are determined by the solution of a set of nonlinear algebraic equations. The solution of these algebraic equations is obtained by the use of a digital computer.

Appendix

since the infinite sum of the $b_{k,j+1}$'s is finite.

Transient Behavior

While it has been shown that $x(k)$ converges to the desired state x^r in the mean square as $k\to\infty$, nothing has been said about the rapidity or lack of it in the convergence. From equation 20 in the absence of noise (i.e., $n(k)=0$), it can be seen that V_k, the variance of the system error, is less than

$$b_{k0}=\prod_{j=0}^{k}\mu_j$$

times the variance of the initial error. Therefore a fast transient response requires that b_{k0} tends to zero quickly with k. In the previous example, with

$$\mu_k\geq\left|1-\frac{A}{1+k}\,Tg(k)\right|^2$$

it is apparent that this result depends on getting μ_k very small (i.e., $\mu_k<1$) as quickly as possible. This depends on $g(k)$ and AT but since only AT is commandable it seems that the best one can do is choose AT so that μ_k is less than unity for all k. (Note that as $k\to\infty$, $\mu_k\to 1$.) This choice depends on the bounds α_1 and α_2 on $g(k)$. However, any particular transient response depends on the exact gain sequence, $g(k)$, and the noise, $n(k)$, so that often another choice of A which leads to μ_k being greater than unity for small k, yields better results.

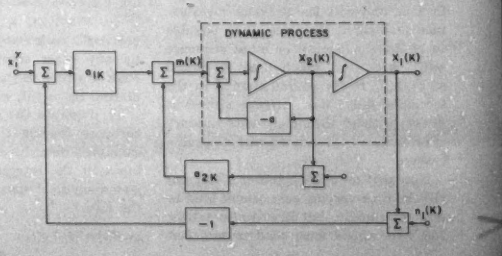


Fig. 6. Example 2

bility in a stochastic sense is conventional except for time-varying gain elements.

It is hoped that the present discussion will help in developing the design of control systems in which the a priori knowledge of the description of the process is meager and measurement is difficult.

References

1. Self Optimizing Systems with Random Inputs, R. G. C. Britt. *Report*, Seminar on Non-Linear Control Problems, Department of Engineering, Cambridge University, Cambridge, England, Sept. 1954, pp. 88–89.

2. Servos That Use Logic Can Optimize, R. L. Cosgriff. *Control Engineering*, New York, N. Y., Sept. 1955, pp. 133–35.

3. A Self-Adjusting System for Optimum Dynamic Performance, G. W. Anderson, J. A. Aseltine. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1958, pp. 182–90.

4. Design of a Self-Optimizing Control System, R. E. Kalman. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 80, 1957, pp. 468–78.

5. A Stochastic Approximation Method, H. Robbins, S. Monro. *Annals of Mathematical Statistics*, East Lansing, Mich., vol. 22, 1951, pp. 400–07.

6. On the Stochastic Approximation Method of Robbins and Monro, J. Wolfowitz. *Ibid.*, vol. 23, 1952, pp. 457–61.

7. Approximation Methods Which Converge with Probability One, J. R. Blum. *Ibid.*, vol. 25, 1954, pp. 737–44.

8. Multidimensional Stochastic Approximation Methods, J. R. Blum. *Ibid.*, pp. 737–44.

9. On Stochastic Approximation, A. Dvoretzky. *Proceedings*, Third Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley, Calif., vol. 1, 1957, pp. 39–51.

10. General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems, R. E. Kalman, J. E. Bertram. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, 1958 (Jan. 1959 section), pp. 602–09.

11. The Approximate Solution of Matrix Problems, A. S. Householder. *Journal*, Association for Computing Machinery, Baltimore, Md., vol. 5, no. 3, July 1958.

12. Theory and Application of Infinite Series (book), Knopp. Hafner Publishing Company, New York, N. Y., 1947, pp. 218–26.

13. Sampled-Data Control Systems (book), J. R. Ragazzini, G. Franklin. McGraw-Hill Book Company, Inc., New York, N. Y., 1958, chap. 10.

The Effect of Elevated Temperature on Flash-Welded Aluminum-Copper Joints

C. R. DIXON
ASSOCIATE MEMBER AIEE

F. G. NELSON
NONMEMBER AIEE

BOTH ALUMINUM and copper have high conductivity and hence are excellently suited for use as electric conductors. Since both metals are widely used, connections between them frequently have to be made. The most common type of connection is the bolted joint, as when an aluminum bus bar is bolted to a copper stud. Although this type of connection is satisfactory, frequently a more efficient and less bulky connection, with better appearance, can be made by flash welding, as illustrated in Fig. 1.

The quality of flash welds is dependent on certain conditions during the welding operation. As a result of important improvements in techniques, especially in the forgoing mechanism and in the die design of the welding equipment, flash welding is now used commercially in an increasing number of applications in which connections have to be made between aluminum and copper. The improvements provide 1. rigid support of the pieces at the ends during joining, 2. large increase of velocity of the platen at the time of upset, and 3. control of time of change from flashing to upsetting during the flash-welding operation. These three improvements make it possible to obtain welds of high quality. Furthermore, the die not only rigidly supports the material but uses the energy of motion to shear off the upset metal.

While it was possible that extreme temperature conditions might be harmful to the joints, it was believed that the conditions that would exist in normal service would not be harmful. Some tests to evaluate the effects of elevated temperatures were reported several years ago, using joints made before the present improved techniques were available.[1] In the investigation now being described, the improved techniques have been used. From tests of almost 500 joints, the effects of heating on tensile strength, ductility, resistance to impact, and electrical resistance were determined.

Welding Procedure and Materials

A general view of the 250-kva flash welder that was used is shown in Fig. 2. A close-up of the dies, indicating the pinching-off action at the weld, is shown in Fig. 3. An oscillogram of the operation during the controlled change of platen movement, when the velocity of the platen changes from 0.125 inch per second to 13.7 inches per second, is shown in Fig. 4, together with traces for voltage and current.

Joints were prepared consisting of commerically pure aluminum bar, 1/4 by $1^1/_2$ inches, flash welded to copper bar of the same size.

The conditions under which the welds were made are shown in Fig. 5. Before making any welds to be tested, the setup of the welding machine was checked by

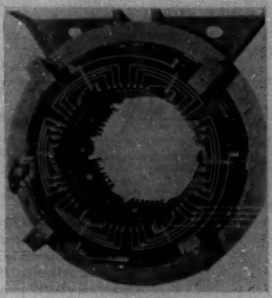


Fig. 1. Stator of large generator containing flash-welded aluminum-copper joints. Arrows indicate one of 24 groups of welds; note bend near one weld

Fig. 2. Flash-welding machine, 250 kva

Paper 59-1162, recommended by the AIEE Electric Welding Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted April 8, 1959; made available for printing August 6, 1959.

C. R. Dixon is with Alcoa Process Development Laboratories, and F. G. Nelson is with Alcoa Research Laboratories, both in New Kensington, Pa.

Fig. 3 (left). Pinch-off dies in flash - welding machine

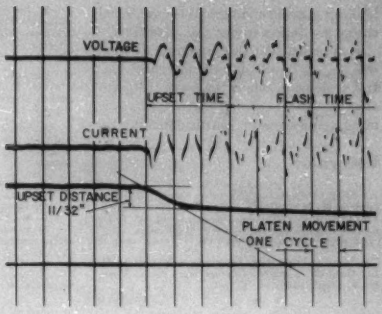


Fig. 4 (right). Oscillogram indicating speed of platen during flash welding

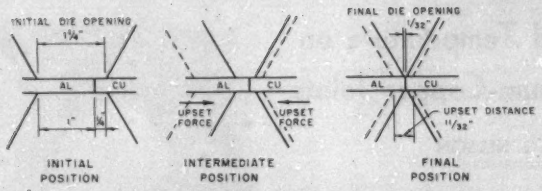


Fig. 5. Conditions used in flash welding 1/4-inch by 1 1/2-inch bars

Flashing distance = 7/8 inch
Flashing time = 7 seconds
Flashing velocity = 0.125 inch per second
Force time delay = 0.032 second
Upset time = 0.049 second
Maximum upset velocity = 13.7 inches per second

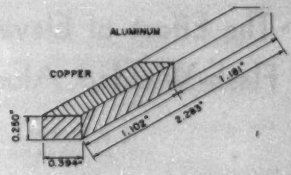
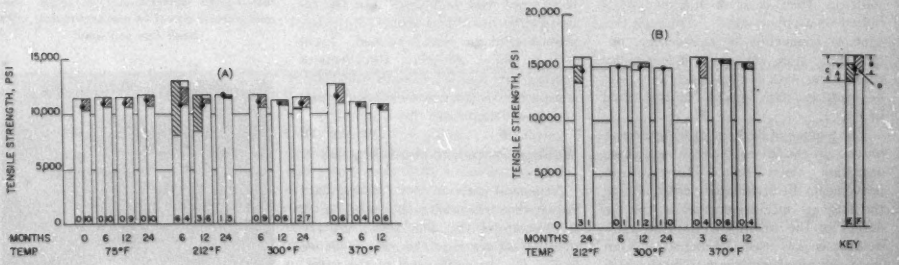


Fig. 6. Cantilever-beam impact test specimen of flash-welded aluminum-copper joint



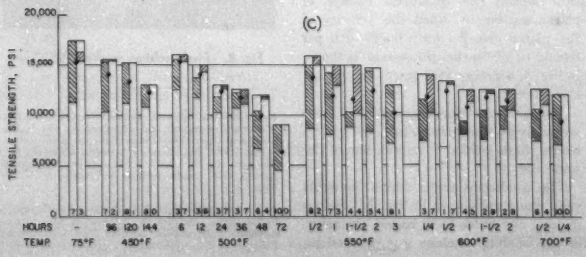


Fig. 7. Tensile strengths at room temperature after indicated exposures for group A (A), group B (B), and group C (C)

Key:
A = range for specimens that failed at weld
B = range for specimens that failed in aluminum
C = range for all specimens
D = average for all specimens
E = number of specimens that failed at weld
F = number of specimens that failed in aluminum

Table I. Average Results of Tensile Tests

Heating Conditions		Tensile Strength, Psi	Location of Failure, No. at Weld per Total No.	Extension in 2 Inches, %*
Temperature, F	Time, Months			
Group A				
75	0	11,000	0/10	
	6	11,200	0/10	
	12	11,200	0/9	
	24	11,500	0/10	
212	6	11,000	8/10	6.9
	12	11,000	3/9	8.7
	24	12,000	1/6	21.0
300	6	11,400	0/9	
	12	11,200	0/6	
	24	11,100	2/9	9.8
370	3†	11,800	0/6	
	6†	11,000	0/4	
	12†	10,900	0/6	
Group B				
212	24	14,800	3/4	7.7
300	6	15,200	0/1	
	12	15,100	1/3	11.5
	24	15,000	1/1	17.0
370	3†	15,500	0/6	
	6†	15,600	0/6	
	12†	15,200	0/4	

Heating Conditions		Tensile Strength, Psi	Location of Failure, No. at Weld per Total No.	Extension in 2 Inches, %*
Temperature, F	Time, Hours			
Group C				
75		15,400	7/10	9.3
450	96	14,000	7/9	10.3
	120	13,300	8/9	5.5
	144	12,200	8/8	5.2
500	6	15,400	3/10	13.5
	12	14,200	3/9	13.5
	24	12,500	3/10	14.2
	36	11,900	3/10	14.5
	48	10,100	6/10	10.9
	72	6,500	10/10	4.0
550	1/2	13,900	8/10	11.9
	1	12,000	7/10	8.6
	1½	11,700	4/8	7.4
	2	12,600	5/9	10.4
	3	10,500	8/9	13.1
600	1/4	12,100	3/10	10.0
	1/2	12,500	1/8	7.6
	1	10,800	4/9	14.9
	1½	11,800	2/10	16.0
	2	11,500	2/10	10.2
700	1/12	10,300	6/10	9.7
	1/4	9,400	10/10	6.6

* Includes only those that failed at weld.
† Intermittent instead of continuous heating; total time at temperature.

qualification tests in which sample welds were required to withstand bending 90 degrees around a 3/4-inch mandrel without failure.

The joints consisted of three groups. Groups *A* and *B* were welded at the same time, the only difference being that the aluminum bar used for group *A* had been annealed while that used for group *B* was in an intermediate strain-hardened temper. The third group, *C*, was welded some 2 years later using bar similar to that used for group *B*.

Test Procedures

Joints from groups *A* and *B*, as shown in Table I, were tested either after holding at room temperature or after heating continuously at temperatures of 212 or 300 (degrees Fahrenheit) for different periods, including some as long as 2 years. In addition, some were heated intermittently at 370 F for a total time at temperature of 1 year. Joints from group *C* were heated continuously at temperatures of 450 to 700 F for periods ranging from a few minutes to 6 days.

All tests were made at room temperature. They included tensile and impact tests, electrical resistance measurements, and metallographic examinations.

Tensile tests were made of specimens consisting of the full section of the bar. As a measure of ductility, the extension, in per cent, of a 2-inch gage length centered across the weld was determined.

Impact test specimens 0.250 by 0.394 inch as in Fig. 6 were machined without notches and were tested as cantilever beams in an Izod machine of 50-foot-pound capacity. The aluminum end of the specimen was gripped in the vise with the joint at the edge of the jaws. The copper end was struck on a 0.394-inch face by the pendulum so that an original surface was subjected to maximum tension.

Table II. Results of Impact Tests

Heating Conditions		Total No. of Tests	Complete Fractures		
Temperature, F	Time		No.	Range of Values, Foot-Pounds	Average Value, Foot-Pounds
Groups A and B, Time in Months					
75	0	9	0		
	6	9	1	2.3	2.3
	12	9	1	3.2	3.2
	24	9	2	6.6–6.6	6.6
212	6	9	2	2.4–3.7	3.0
	12	9	1	0.9	0.9
	24	9	1	1.6	1.6
300	6	9	2	2.9–7.7	5.3
	12	9	0		
	24	9	2	3.4–8.0	5.2
370	3*	9	3	3.7–12.1	7.2
	6*	9	0		
	12*	9	1	1.6	1.6
Group C, Time in Hours					
75		4	2	1.5–8.0	4.8
450	96	9	5	0.5–6.0	2.4
	120	9	9	0.5–7.6	2.3
	144	9	5	0.3–2.5	1.3
500	6	9	3	2.1–2.5	2.2
	12	9	2	1.3–4.1	2.7
	24	9	2	2.5–3.3	2.9
	36	9	1	1.1	1.1
	48	9	3	1.0–3.5	4.1
	72	9	6	1.0–4.5	2.9
550	1/2	9	1	4.8	4.8
	1	9	0		
	1½	9	2	0.5–2.5	1.5
	2	9	1	1.5	1.5
	3	9	1	1.0	1.0
600	1/4	9	2	2.5–5.5	4.0
	1/2	9	1	0.5	0.5
	1	9	1	2.2	2.2
	1½	9	0		
	2	9	2	0.2–1.8	1.0
700	1/12	9	3	0.8–2.0	1.6
	1/4	9	5	1.0–6.1	2.8

* Intermittent instead of continuous heating; total time at temperature.

Electrical resistance measurements were made in which current was passed through the length of the specimen, including the layer of intermetallic compound. The potential drops were measured over 2- and 4-inch gage lengths, using an L-and-N-type *K-2* potentiometer.

Metallographic examination of some

Table III. Electrical Resistance of 1/4- by 1 1/2-inch Flash-Welded Aluminum-Copper Joints

Heating Conditions		Electrical Resistance, Microhms, Gage Length	
Temperature, F	Time	4 Inches	2 Inches
Groups A and B, Time in Months			
Room temperature		10.0	5.0
212	24	10.1	5.0
300	24	9.9	4.8
370	12*	10.0	
Group C, Time in Hours			
Room temperature		9.8	4.8
450	144	9.7	4.8
500	72	9.8	5.0
550	3	9.8	4.8
600	2	9.7	4.7
700	1/4	9.8	4.8

* Intermittent instead of continuous heating; total time at temperature.

of the welds was made in which the thickness of the layer of intermetallic compound was determined.

Results and Discussion

The results of tensile tests are summarized in Table I and in Fig. 7.

In about 40% of the tests, failure occurred in the aluminum and, therefore, the actual strengths of the welds were not determined, only the maximum stress that they withstood before the specimens failed. The average strength of all of the specimens was 11,900 psi (pounds per square inch), but for the reason just stated the actual average strength of the welds must have been even higher.

In groups *A* and *B* (Table I and Figs. 7(A) and (B)) there was obviously no significant decrease in the strength of the specimens as a result of heating. It will be noted that in both groups the slight decreases in strengths with increases in length of heating period at the highest temperature (370 F) are not an indication of loss of strength of the welds, since none of the failures in these tests occurred at welds. Since very few of the specimens in groups *A* and *B* failed at the welds, the higher strengths in group *B* reflect the higher strength of the aluminum bar in those specimens.

In group *C* (Table I and Fig. 7(C)), the percentage of failures at the welds was larger than in groups *A* and *B*. This is true regardless of whether or not the joints were heated. Also, heating at the various conditions caused decreases in strengths of the specimens. Decreases of this kind were expected because such heating would anneal the aluminum which, in this group of specimens as mentioned previously, was in a strain-hardened temper. When the decreases that occurred were accompanied by definite increases in the proportion of failures at the welds, and when the strengths became less than that of thoroughly annealed aluminum, then there was distinct evidence that the heating was detrimental to the welds. For the heating conditions investigated, the percentage of failures at the weld did not increase significantly until the specimens had been heated 120 hours at 450 F, 48 hours at 500 F or 1/4 hour at 700 F. The average tensile strength did not decrease to less than 10,000 psi until the specimens had been heated 72 hours at 500 F or 1/4 hour at 700 F. In general, the scatter or spread of strengths was no greater for the exposed specimens than for those that were not heated, but the lowest individual strengths encountered for exposures of 48 hours or more at 500 F and for all times at higher temperatures were considerably below 10,000 psi.

In general, the extension values obtained in the tensile tests are in agreement with the foregoing observations.

The results of the impact tests are summarized in Table II. Values are shown only for those tests in which complete fracture occurred, because it is only under such conditions that the results have quantitative meaning.

Complete fracture occurred in only about one-fourth of the tests, and in those cases it was at the weld. Generally, the percentage of complete fractures tended to decrease with increase in temperature. Although the average impact values were generally lower for the specimens that had been heated than for the specimens that had not been heated, the number of specimens that actually

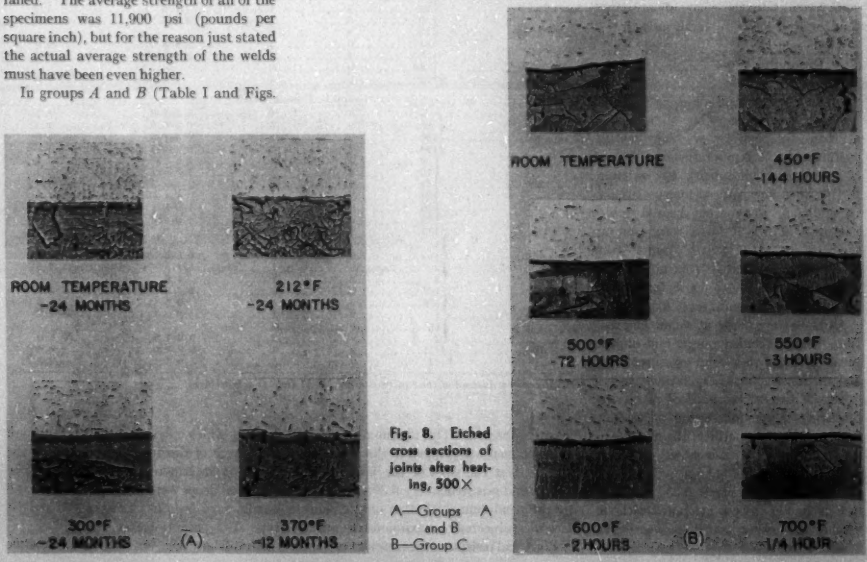


Fig. 8. Etched cross sections of joints after heating, 500×

A—Groups A and B

B—Group C

fractured at the welds was so small, and the number of specimens in which the welds withstood the impact blow was so large, that the values themselves are not significant in evaluating the effects of heating.

The measurements of electrical resistance at room temperature after heating for the longest period of time at each temperature are listed in Table III. The data show no trend toward an increase in resistance with increase in temperature or time of heating. Furthermore, the resistance of a 2-inch gage length was almost exactly half that of a 4-inch gage length. If the thickness of the layer of intermetallic compound had been increased by the heating, the resistances would have increased and the resistance for a 2-inch gage length would have been more than half of those for a 4-inch gage length. This would also be the case if any discontinuities, such as separation at the intermetallic layer, had been caused by heating.

Photomicrographs of welds subjected to each of the temperatures for the longest periods are shown in Fig. 8 There is little if any significant increase in thickness of the interfacial layer with increase in temperature or length of heating period. The thickness of this layer generally was 0.00015 inch or less.

Conclusions

Improved techniques have been developed by which flash-welded joints between aluminum and copper can now be made commercially. The average strengths of the welds are at least 10,000 psi when heated as high as 300 F for 2 years, 370 F for 1 year, 450 F for 144 hours, 500 F for 36 hours, 600 F for 2 hours, or 700 F for 5 minutes. The strengths of individual welds, however, may drop below 10,000 psi when joints are heated 48 hours or more at 500 F, or when heated for shorter times above that temperature.

There was no evidence of embrittlement at any of the heating conditions used.

Reference

1. Preventing Embrittlement in Copper-to-Aluminum Weld Joints, C. L. Carlson, R. M. Leedy. *Product Engineering*, New York, N. Y., Oct. 1955, pp. 172–73.

Discussion

Thomas B. Correy (Hanford Atomic Products Operation, General Electric Company, Richland, Wash.): The subject matter is very well prepared and the duration of the tests long enough to provide accurate vital information on the joining of aluminum to copper by the flash-welding process. This information is very vital as portions of the electrical industry are moving rapidly into the use of aluminum for windings in motors, generators, and transformers.

There is one portion of the tensile test data that the authors neglected to discuss. What causes the high weld failure rate for the 212 F test? Other alloys of aluminum have displayed brittle failure of flash-welded joints at a critical temperature somewhat higher than this. This condition should be explained, as windings are subject to temperatures of this order of magnitude and under mechanical stress.

C. R. Dixon and F. G. Nelson: The authors thank Mr. Correy for his complimentary remarks and admit that he has asked a very pertinent question concerning the data presented.

Efforts were made to determine the reason for the relatively high number of failures in the welds of group *A* that were subjected to a temperature of 212 F. The fractured specimens were examined and both the fractured surfaces and the metallurgical structure appeared normal. The only conclusion the authors can reach is that this was simply the result of random selection of specimens. Experience with similar flash-welded joints in current-carrying circuits in Alcoa plants where they have been in use at temperatures near 212 F has given us confidence in the behavior of flash-welded joints at this temperature.

Transient Response and the Stabilization of Feedback Amplifiers

J. H. MULLIGAN, JR.
MEMBER AIEE

ANY VACUUM-tube or transistor amplifier is continuously subjected to random current and voltage disturbances of varying peak amplitude and waveshape. Thermal noise is a common example of such excitation. The importance of this interfering effect, as far as the circuit designer is concerned, depends to a large extent upon the relative magnitudes of the desired signals for which the amplifier is designed as compared to the strength of the response produced by the system to the random excitation.

In a passive network or a nonfeedback amplifier, the transient response to an impulse or step function consists of constants and/or exponentially damped terms, and, as a consequence, one may predict accurately the response of any applied signal be it desired or undesired. In a feedback amplifier, on the other hand, possibilities exist, in general, for the production of terms in the transient response whose amplitudes continue to increase with time until such values are reached that the nonlinear input-output characteristics of the vacuum tubes or transistors limit further growth. At such times, it is possible for the response to bear little relation to the excitation which initiated it. Because of the continual excitation by thermal noise pulses and similar internally generated signals, an amplifier with no external signal applied can produce an appreciable output with no apparent input because of the presence in its transient response of terms which increase exponentially with time.

Amplifiers and other dynamic systems can accordingly be classified as stable or unstable, respectively, depending on the nature of their transient response. In this paper a stable system is considered to be one in which any disturbance which decays with time and ultimately dies out produces a response which decays with time and ultimately dies out. The principal concern here is with the process of stabilizing an unstable feedback amplifier. Stated in other terms, the problem considered is one of determining what modifications are needed in a feedback amplifier design in order to change the

Paper 59-1151, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted March 23, 1959; made available for printing July 31, 1959.

J. H. Mulligan, Jr., is with New York University, New York, N. Y.

transient response so that it contains no terms whose amplitudes increase indefinitely with time. In terms of the complex frequency plane, the stabilization problem consists of insuring that all roots of the characteristic equation of the amplifier lie in the interior of the left half plane.

The stabilization problem for feedback amplifiers is not a new one by any means. The question has been considered at length in terms of Nyquist polar diagrams as well as in terms of plots of the gain and phase characteristics of the return ratio, or loop gain, as functions of frequency. H. W. Bode[1] has given an elegant solution to the stabilization problem in terms of an "ideal gain characteristic." This loop gain function calls for a constant gain up to a certain frequency, and a constant phase shift equal to 180 degrees minus a prescribed "phase margin" beyond, until the gain has dropped a prescribed amount (called the "gain margin") below unity. Design control is relaxed above the frequency at which this occurs. The characteristic is "ideal" in the sense that it achieves the maximum amount of feedback over a prescribed frequency range that can be attained with the given tubes or transistors in an absolutely stable amplifier whose phase shift does not exceed a fixed amount in the frequency interval in which design control is maintained.

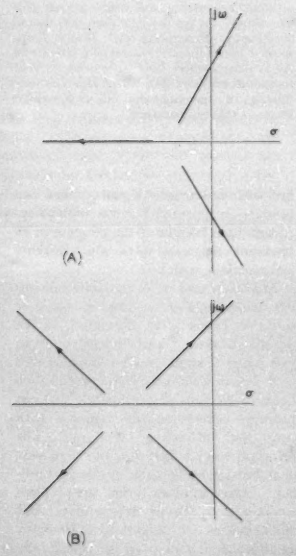


Fig. 1. Asymptotes approached by loci of closed loop poles as mid-band open loop gain is increased indefinitely

A—For an amplifier with three more poles than zeros in T(s)

B—For an amplifier with four more poles than zeros in T(s)

Duerdoth[2] has shown, however, that if one relaxes the Bode condition of constant phase shift in the region just outside the pass-band, the loop gain attainable over the useful frequency band can be increased and adequate protection against oscillation which might result from component variation can still be provided. Other investigators have considered the stabilization problem, also from the frequency response viewpoint, with various special design requirements in mind such as the one of obtaining a maximal flatness closed-loop gain response.

The two quantities, gain margin and phase margin, are included in all the approaches to the stabilization problem based on the frequency response viewpoint. Some success has been attained in relating frequency response data to certain aspects of the transient response.[3,4] Gain margins and phase margins are still established to a large extent, however, on a rule-of-thumb basis in the formulation of design requirements rather than being amenable to a direct analytical determination from a knowledge of the transient performance desired from the amplifier to be designed.

It is recalled that the stabilization problem arises in the first place because the transient response is not satisfactory, viz. not all of the roots of the characteristic equation lie in the left half plane. It is profitable, therefore, to consider the stabilization of feedback amplifiers directly in terms of the transient response desired. That is to say, it is of value to study the problem without recourse to frequency response data and, in particular, to think in terms of a performance criterion for stabilization different from a gain margin plus a phase margin. A similar approach has been utilized in connection with the design of servomechanisms, but the results obtained cannot be applied directly to the general feedback amplifier problem because of a fundamental difference in the elements which find principal application in the two classes of systems.

In servomechanisms, the time constants, or frequency response characteristics, associated with the mechanical elements employed represent a fundamental limitation to the speed of system response. The limitations of vacuum tubes and transistors which arise from parasitic capacitances and the variation of parameters with frequency are of negligible importance when these elements are employed in servomechanism systems. This occurs because the time constants associated with these devices are so much less than those associated with the response of the necessary mechanical elements that their limitations are virtually nonexistent by comparison. On the other hand, the so-called asymptotic characteristics of tubes and transistors (their limiting performance at high frequencies) impose serious constraints on pole and zero locations in the complex plane that may be realized in feedback or nonfeedback amplifiers. Taking these extra constraints into account is an important feature of a design procedure for the feedback amplifier problem.

The principal purpose of this paper is to describe an approach to the stabilization of feedback amplifiers which is based on the direct control of the transient response to a step function input. The minimum rate at which the overshoots at succeeding maxima of the step function response are attenuated is taken as a design criterion for stabilization and the direct determination of element values from the specified attenuation is shown in illustrative examples.

Stabilization Philosophy Based on Transient Response

Root Loci and Dominant Response

In the stabilization philosophy based on the direct control of transient response, the requirements calling for a given phase margin and a preassigned gain margin are replaced by either a statement of the minimum rate at which the overshoots at the succeeding maxima of the step-function response are to be attenuated or, alternatively, a specification of the maximum per-cent overshoot that will be permitted in the step-function response. For either one of these two time-response design criteria to have practical usefulness, one must have reason to believe that the step-function response characteristics of common feedback amplifiers are sufficiently similar so that the proposed criteria are both meaningful and applicable.

Two factors considered together indicate that the transient response approach to the stabilization problem should be valuable in the design of a large class of vacuum-tube and transistor feedback amplifiers. The first factor is concerned with the location of the poles of the

closed-loop transfer function of a feedback amplifier. Root locus techniques introduced by Evans[3] predict an asymptotic movement of the closed-loop poles along radial lines having equal angular displacement and a center located in the left half plane as the mid-band return ratio (or loop gain) is increased without limit. The detailed form of the loci depends on the specific positions of the poles and zeros of the return ratio function $T(s)$, but the asymptotic behavior for large values of $T(0)$ has the general form depicted in Fig. 1. The movement of the poles away from one another with increased $T(0)$ is suggested from the figure.

This behavior suggests the second factor of importance, namely the applicability of the dominant-term approximation to the step function response introduced previously by the author.[4] In the reference cited, it was shown that under suitable conditions of separation among the poles in the complex frequency plane, the normalized time response can be well approximated at the first maximum and beyond by a constant term plus a single exponentially damped sinusoid. In general, although not always, the pair of complex poles located nearest to the imaginary axis in the complex frequency plane will produce the dominant term in the time response. The approximation to the actual time response given by the dominant term theory is improved as the separation among poles increases. By consideration of the root loci asymptotes obtained from the open-loop transfer functions of common vacuum-tube and transistor feedback amplifiers and the corresponding pole movements which are implied, one is led to the conclusion that the dominant term theory should have wide application in the approximation of the transient response of such amplifiers. A design criterion for stabilization based on the form of time response obtained using the dominant-term approximation should, therefore, have considerable utility.

Underlying Principles

For a dynamic system having a transfer function consisting of a single pair of complex poles as depicted in Fig. 2, the first maximum in the time response to a step function excitation occurs at a value t_{max} expressed by

$$\beta_1 t_{max} = \pi \tag{1}$$

and the deviation (γ) from unity at the first maximum of the normalized time response is[5]

$$\gamma = e^{-\pi \cot \psi_1} \tag{2}$$

The quantities β_1 and ψ_1 are shown in Fig. 2. For a pole-zero configuration to which the dominant term theory is applicable (detailed conditions relating to the degree of approximation involved are given in reference 6) but which is otherwise arbitrary, the extreme values of the time response, local maxima and minima, occur at times given approximately by

$$\beta_1 t = h\pi - \lambda_1 \tag{3}$$

and the corresponding deviations from unity in the normalized response are approximately given by equation 4. (See equations 23 and 40 of reference 6 for the approximations associated with equations 3 and 4.)

$$\gamma_h = (-1)^{w+h+1} M_1 e^{\lambda_1 \cot \psi_1} e^{-h\pi \cot \psi_1} \tag{4}$$

where w is the number of real zeros in the right half plane and $h = b, b\text{-}1, b\text{-}2, \ldots, b$ being the integer (positive, negative, or zero) for which $b\pi - \lambda_1$ has its smallest non-negative value. M_1 is the coefficient of the term $\sin(\beta_1 t + \lambda_1 + \psi_1)$ in the normalized time response. It is equal to the quotient of the product of length factors due to zeros divided by the product of length factors due to poles. All poles and zeros in the pattern except the pole at $-\alpha_1 - j\beta_1$ are considered; the factor for any element, pole or zero, is equal to the ratio of the distance in the complex plane from the element to $-\alpha_1 + j\beta_1$ divided by the distance from the origin to the element. Each directed distance from a pole or zero to $-\alpha_1 + j\beta_1$ may be viewed as a complex number. λ_1 is found by summing the angles of the complex numbers associated with the directed distance from each pole and zero to $-\alpha_1 + j\beta_1$, a plus sign being used with angles associated with zeros and a minus sign being employed for the angles associated with poles. The pole at $-\alpha_1 - j\beta_1$ is not included in this calculation.

By examination of equation 4, it is seen that the ratio of the deviations from unity in the normalized time response to two successive extreme values is

$$\eta = \left| \frac{\gamma_{h+1}}{\gamma_h} \right| = e^{-\pi \cot \psi_1} \tag{5}$$

If there are no zeros in the right half plane and $b = 1$, which is common for simple pole-zero patterns, then the first maximum occurs at t_{max} where

$$\beta_1 t_{max} = \pi - \lambda_1 \tag{6}$$

and the fractional overshoot at the first maximum is

$$\gamma_1 = M_1 e^{\lambda_1 \cot \psi_1} e^{-\pi \cot \psi_1} \tag{7}$$

The quantity $M_1 e^{\lambda_1 \cot \psi_1}$ is a function not only of the pole angle of the dominant pair of poles but also of the location of the other poles and zeros relative to $-\alpha_1 \pm j\beta_1$. It has been defined as an "overshoot factor," and graphs have been provided for its direct computation from pole-zero locations.[6] It is noted by comparison of equations 1 and 6 as well as equations 2 and 7 that in systems where the dominant-term theory applies, the effect of adding other poles and zeros to the simple system of two complex poles is but to change both the magnitude and the time of occurrence of the first maximum, leaving the shape of the curve essentially unchanged. These matters are considered at length in the reference cited.

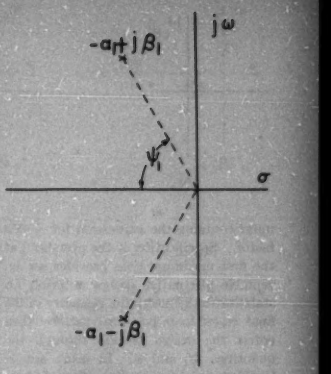


Fig. 2. Pole pattern for a system having a transfer function containing two complex poles illustrating notation used in text

Stabilization Design Criterion

Equation 5 provides a logical design criterion for the stabilization of feedback amplifiers. For a system to be stable it is necessary that $\eta < 1$, or that $\psi_1 < \pi/2$. Furthermore, the time taken for the step-function transient to be completed to an arbitrarily assigned degree; i.e., the time taken for the value of overshoot at a local maximum to become less than a stated value, is not only a measure of the stability of the amplifier but one which is acceptable from an intuitive standpoint. η will be specified, therefore, as a stabilization design criterion based on the control of transient response; a design philosophy will be outlined whose use will permit achieving a pair of dominant poles with the angle ψ_1 once the angle has been determined from an acceptable value of η. It is to be noted also, however, that equation 7 for the overshoot at the first maxi-

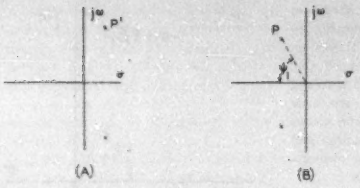


Fig. 3. The feedback amplifier stability problem may be viewed as that of shifting the position of the complex pole at P' (and its conjugate) in (A) to point P in (B)

mum contains the expression for η as a factor. Specification of the overshoot at the first maximum thus provides an acceptable alternative design criterion for stabilization although the geometry of the final closed-loop pole-zero pattern then enters the design process through the quantities M_1 and λ_1. In many simple feedback amplifier configurations (where the excess of poles over zeros is three or four, for example), the overshoot factor $M_1 e^{\lambda_1 \pi} \cot \psi_1$ is frequently approximately unity so that η is essentially equal to γ_1, the percent overshoot at the first maximum.

A statement of the stabilization problem can now be made in terms of the complex frequency plane. Suppose that an amplifier has been designed to meet various performance requirements and it is found to have two complex poles in the right half plane as depicted in Fig. 3(A). Furthermore, suppose the circuit designer desires that the step-function transient response should have a maximum overshoot of a prescribed percent or that the damped sinusoidal dominant transient should decay at a certain minimum rate. In effect, the designer wishes to achieve the condition pictured in Fig. 3(B), a condition in which the dominant pole pair (P in the second quadrant) has an angle ψ_1 determined from either the overshoot or attenuation requirement. (Note that the two are related; see equations 5 and 7.) The amplifier design must be altered in such a manner that with the desired mid-band return ratio unchanged, point P' in Fig. 3(A) moves to P in Fig. 3(B). To illustrate the process involved in the shift of dominant pole position, let the closed-loop transfer function be written in the form of equation 8. (This can be done only for a certain class of feedback amplifiers.)

$$G(s) = k \frac{T(s)}{1+T(s)} \tag{8}$$

where $T(s)$ is the return ratio for zero reference for one of the active elements in the amplifier and k is independent of s. The poles of $G(s)$ are located at those values of s which result in the equality $T(s) = -1$. In order to cause the root locus to pass through point P, it is necessary that at point P

$$|T(s)| = 1 \tag{9}$$

$$\text{Angle } T(s) = \pm(2n-1)\pi, \quad n = 0, 1, 2, \ldots \tag{10}$$

In summary, the stabilization procedure based on direct control of transient response consists of choosing ψ_1 on the basis of overshoot at the first maximums or on the basis of the rate of attenuation of the transient, plus the subsequent imposition of the conditions 9 and 10 at a point on the radial line through the origin inclined ψ_1 degrees to the horizontal, Fig. 3(B). As will be evident, additional constraints can be also introduced into the design process, if desired.

The actual technique employed to effect the variation in dominant pole position depends on the specific amplifier design problem. Two practical techniques that can be employed for the purpose are the addition of corrective interstage networks and the use of local feedback around the active elements. Both methods are illustrated in the examples furnished.

Implementation

The statement of the stabilization problem in the form previously mentioned is closely related to the problem of designing a servomechanism for a prescribed transient response, a problem for which an important body of design information exists. The contributions of Aaron, Aseltine, Chu, Russell and Weaver, and Zaborsky[7-11] are representative of the progress that has been made in this area. And yet, the feedback amplifier stabilization problem differs from the former in one crucial respect which was previously mentioned. When vacuum tubes and transistors are used as elements of feedback amplifiers, their limiting behavior at high frequencies or "asymptotic characteristics" are of paramount importance in the design process. When they are used as elements in a servomechanism system, however, the limitations of the mechanical elements virtually obscure their limitations. As a result, the addition of amplifier stages with suitable interstage networks to a servomechanism system provides the circuit designer with a considerable amount of flexibility in the realization of suitable pole-zero configurations to control transient response. In the low-frequency spectrum associated with servomechanism operation, the limit of amplification that may be added occurs primarily as a result of the undesirable effects of overloading amplifier stages or the excessive amplification of noise. In the extended frequency range in which feedback amplifiers find their usual application, however, the constraints on the pole-zero locations imposed by the properties of a tube or transistor itself are usually important design factors. In general, in terms of the active devices now available, the addition of one or more stages of amplification to an existing feedback amplifier whose return ratio characteristics are satisfactory except for the problem of stabilization and transient response can do more to complicate the design problem than to simplify it because of the additional constraints that are imposed.

If one measures the frequency response of a grounded cathode vacuum tube

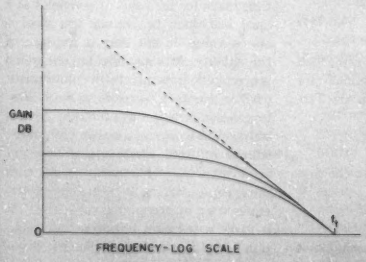


Fig. 4. Gain characteristic of single-stage grounded cathode amplifier showing approach to asymptotic characteristic (dashed line-slope: −6 db per octave). Figure of merit frequency $f_t = g_m/2\pi C$. Increased load resistance yields higher low-frequency gain, less bandwidth, but the same asymptotic behavior at high frequencies

Fig. 5. Simple resistance-capacitance interstage for illustrative example

amplifier for various values of load resistance, curves similar to those shown in Fig. 4 will result. As frequency is increased, the curves for all values of load resistance become asymptotic to the dashed line which has a slope of -6 db (decibels) per octave and crosses the zero gain level at the frequency $g_m/2\pi C$. The location of the dashed line, called the "asymptotic characteristic," is clearly dependent upon the properties of the tube and the minimum shunt capacitance associated with it. Comparable behavior is experienced in transistor amplifiers. Asymptotic characteristics are of fundamental importance in the stabilization procedures based on frequency response.[1] A convenient classification based on these characteristics may be made of the various techniques used to shift the root locus so that it passes through P in Fig. 3(B) rather than through P' in Fig. 3(A) for the same mid-band return ratio. The techniques may be grouped into two classes; viz. those that do not affect the asymptotic characteristic and those that degrade it. Those in the first class permit optimization of design with respect to selected parameters whereas the techniques in the second class usually permit simplifications in design or adjustment at the expense of not being able to attain the ultimate in performance. Simple examples of both methods are given in the next section.

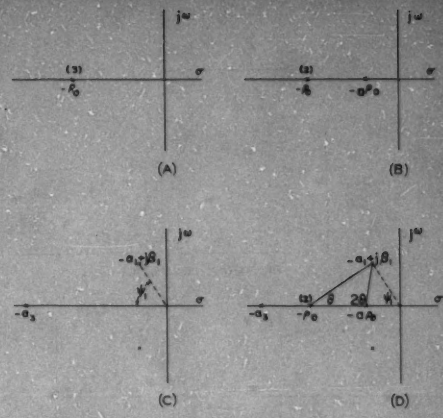


Fig. 6. Pole patterns for illustrative example

A—Original pole pattern of $T(s)$
B—Modified pole pattern caused by addition of capacitor to one stage
C—Desired roots of $1+T(s)=0$
D—Illustration of variables used in formulating design conditions

Illustrative Examples

In this section, three examples are considered which illustrate the use of the approach to stabilization based on time response which has been discussed. The examples have been chosen primarily for their value in illustrating the stabilization design philosophy under consideration although all are useful in practical applications.

Shunt Capacitor-Shifted Asymptote

Consider that a 3-stage feedback amplifier having a gain given by equation 8 has been designed with interstage networks having the form shown in Fig. 5, and a return ratio function expressed by

$$T(s)=\frac{T(0)}{\left(1+\frac{s}{\rho_0}\right)^3} \tag{11}$$

with $T(0)=40$. The pole pattern for the function $T(s)$ is shown in Fig. 6(A). It is found that for $T(0)>8$, the function $1+T(s)$ has two complex roots in the right half plane, so the feedback amplifier is unstable. Assume that it is desired to stabilize the amplifier so that $\eta\leq 0.2$, that is, the deviation from unity at one local extreme value of the transient response is to be (at least) five times the deviation at the next maximum deviation. Furthermore, it is assumed that the stabilization is to be effected by the addition of a capacitor of value C_1 in parallel with only one of the interstage networks shown in Fig. 5.

The stabilization method selected has the property of shifting the asymptote of the stage to which the capacitance is added to the left in Fig. 4. The new intersection of the asymptote for the single stage with the zero-db gain level will be at a frequency $f_1=g_m/2\pi(C+C_1)$; the composite asymptote for the whole amplifier will also be shifted to the left resulting in degradation of performance as compared to that attainable from the original structure.

The addition of capacitor C_1 as described produces the pole pattern for the modified return ratio function $T'(s)$ shown in Fig. 6(B) where

$$a\rho_0=\frac{1}{R(C+C_1)} \tag{12}$$

and

$$T'(s)=\frac{T(0)}{\left(1+\frac{s}{\rho_0}\right)^2\left(1+\frac{s}{a\rho_0}\right)}=\frac{a\rho_0^3T(0)}{(s+\rho_0)^2(s+a\rho_0)} \tag{13}$$

Fig. 6(C) shows the pole pattern of $G(s)$ (see equation 8), that is, the root pattern of $1+T(s)=0$ which is derived from Fig. 6(B) and meets the condition that P lies on the radial line with displacement ψ_1 from the negative real axis. By the use of equation 5 with $\eta=0.2$, ψ_1 is found to be 62.9 degrees. In terms of the quantities defined in Fig. 6(D), in order to have the conditions

$$|T'(s)|=1$$

and

angle of $T'(s)=-180$ degrees

satisfied as the root locus crosses the radial line at P, it is necessary that

$$\sin(2\theta-\psi_1)\sin^2(\theta+\psi_1)=\frac{\sin^2\psi_1}{T(0)} \tag{14}$$

After finding θ using equation 14, the quantity a is determined from the expression

$$a=\frac{1}{2T(0)\cos\theta}\,\frac{\sin^2\psi_1}{\sin^2(\theta+\psi_1)} \tag{15}$$

Knowing R, C, and a, the value of C_1 necessary to effect the desired stabilization is computed using equation 12.

It is of interest to record that

$$\alpha_1=\rho_0\frac{\sin\theta\cos\psi_1}{\sin(\theta+\psi_1)} \tag{16}$$

$$\beta_1=\rho_0\frac{\sin\theta\sin\psi_1}{\sin(\theta+\psi_1)} \tag{17}$$

and

$$\alpha_3=2\rho_0\left[1+\frac{a}{2}-\frac{\sin\theta\cos\psi_1}{\sin(\theta+\psi_1)}\right] \tag{18}$$

For $\psi_1=62.9$ degrees, and $T(0)=40$, it is found that $a=0.0105$, $\alpha_1=0.242\ \rho_0$, $\beta_1=0.473\ \rho_0$, and $\alpha_3=1.526\ \rho_0$. Fig. 6(C) is drawn to scale to show the correct relations among α_1, β_1, and α_3.

Shunt Resistance Capacitance Network—Original Asymptote

The stabilization of the amplifier characterized by equation 11 will now be

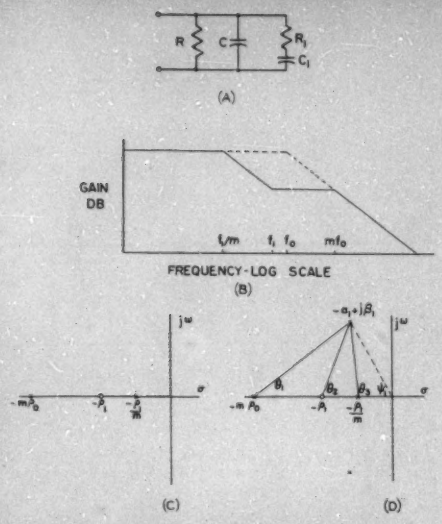


Fig. 7. Stabilization by modification of interstage network

A—Interstage network showing added stabilization elements R_1, C_1

B—Representative gain characteristic of single-stage amplifier using network in (A). The slope of the inclined portions of the characteristic is −6 db per octave; $f_1=\rho_1/2\pi$; see text

C—Pole pattern of $T(s)$ using networks of (A) in three stages. Each pole and zero shown is of third order

D—Illustration of variables used in formulating design conditions

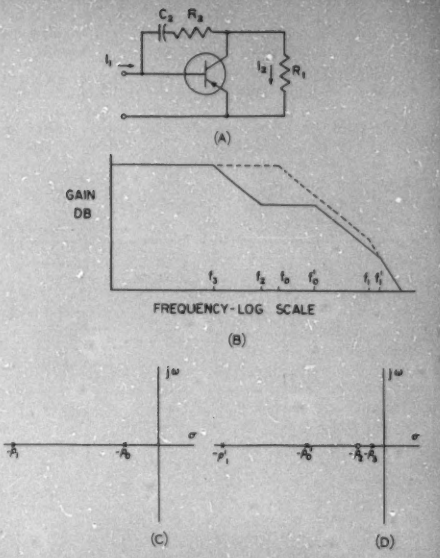


Fig. 8. Stabilization by use of local feedback

A—Common emitter transistor amplifier with local feedback

B—Representative current gain (I_2/I_1) characteristic of stage in (A). The slope of the inclined portions of the characteristic below f_1 is −6 db per octave; above f_1' the slope is −12 db per octave; $f_1=\rho_1/2\pi$; see text

C—Pole pattern of amplifier stage without feedback

D—Pole pattern of amplifier stage with feedback; see text for values of pole and zero locations in terms of circuit elements

undertaken with the same attenuation requirement of $\eta=0.2$ but with the added stipulation that the corrective method employed must not shift the original asymptote of the amplifier.

The addition of a series combination R_1-C_1 in parallel with the interstage of Fig. 5 yields the network of Fig. 7(A). The straight line approximation to the resulting gain characteristic of a vacuum-tube amplifier using the interstage is indicated in Fig. 7(B) by the solid line; the dashed portion illustrates the variation without R_1-C_1. The slope of the characteristic is −6 db per octave. Note that the original asymptote is followed above the frequency mf_0. The form of constraints frequently encountered in the stabilization of amplifiers is also indicated in Fig. 7(B). Note that whereas f_1 may be selected arbitrarily, the other two critical frequencies are determined as soon as the value of m is chosen. This behavior must, of course, be considered in the design process and represents one additional complication not present in the corresponding servomechanism transient response design problem.

The stabilization will be effected by the addition of one R_1-C_1 network to each of the three interstages represented by the characteristic of equation 11. The modified pole-zero pattern for $T(s)$ is shown in Fig. 7(C); all poles and zeros shown are third order. For convenience in the example, m and ρ_1 have been taken the same for all networks. Once ρ_1 and m have been determined as will be discussed, R_1 and C_1 are given by

$$R_1=\frac{R}{m-1}\,\frac{1}{1-\dfrac{\rho_1}{m\rho_0}} \tag{19}$$

$$C_1=\frac{1}{R_1\rho_1} \tag{20}$$

The return ratio with the corrective networks added is

$$T'(s)=\frac{T(0)}{\left(1+\dfrac{s}{m\rho_0}\right)^3}\,\frac{\left(1+\dfrac{s}{\rho_1}\right)^3}{\left(1+\dfrac{ms}{\rho_1}\right)^3}$$

or

$$T'(s)=\frac{\rho_0{}^3T(0)}{(s+m\rho_0)^3}\,\frac{(s+\rho_1)^3}{(s+\rho_1/m)^3} \tag{21}$$

The stabilization problem consists in determining what values to assign to ρ_1 and m so that the root locus passes through the point P at $-\alpha_1+j\beta_1$ located on the dashed line making the angle ψ_1 with the negative real axis when the return ratio has the value $T(0)$. In the preceding example, specification of ψ_1 determined the quantity a uniquely. Here it will be seen that unless another design condition is imposed, there is not just a single pair of values of ρ_1 and m that satisfy the condition imposed on point P but rather a family of pairs. The choice of a particular combination of ρ_1 and m will depend on what other features of the transient response are of interest to the designer.

In order to satisfy the condition $|T'(s)|=1$ at $-\alpha_1+j\beta_1$ in Fig. 7(D), it is nec-

essary that θ_1, θ_2, and θ_3 satisfy the relation

$$\frac{\sin(\psi_1+\theta_1)\sin(\psi_1+\theta_2)}{\sin(\psi_1+\theta_3)}=\frac{\sin\psi_1}{\sqrt[3]{T(0)}} \tag{22}$$

For the angle of $T(s)$ to be an odd multiple of 180 degrees, the three angles must be related as

$$\theta_1-\theta_2+\theta_3=\pm(2n+1)60 \tag{23}$$

As stated, another condition may be imposed on the variables θ_1, θ_2, and θ_3. Assuming the additional condition to be independent of equations 22 and 23, its specification will permit a solution for the three quantities. With θ_1, θ_2, and θ_3 known, the quantities m and ρ_1 from which the necessary values of R_1 and C_1 are computed (see equations 19 and 20) may be calculated from the relations

$$m=\sqrt[3]{T(0)}\,\frac{\sin(\theta_1+\psi_1)}{\sin\psi_1}\,\frac{\sin\theta_2}{\sin\theta_1} \tag{24}$$

$$\rho_1=m\rho_0\,\frac{\sin\theta_1}{\sin(\psi_1+\theta_1)}\,\frac{\sin(\psi_1+\theta_3)}{\sin\theta_3} \tag{25}$$

The co-ordinates of the dominant pole are also of interest. They may be computed from the equations

$$\alpha_1=m\rho_0\,\frac{\sin\theta_1\cos\psi_1}{\sin(\psi_1+\theta_1)} \tag{26}$$

$$\beta_1=\rho_0\sqrt[3]{T(0)}\,\frac{\sin\theta_1\sin\theta_2}{\sin\theta_3} \tag{27}$$

If no specific condition is added to equations 22 and 23, one may choose θ_1 using as a guide, for example, some general characteristics of the transient response. Recall that the time at which the first maximum occurs is inversely proportional to β_1 if λ_1 remains fixed. One might want to select a value of θ_1 based on making β_1 as large as possible. If θ_1 is chosen, then θ_3 may be determined at once from the relation

$$\tan\theta_3=-\frac{\sin\psi_1-A\sin(\psi_1+\theta_1-60^\circ)}{\cos\psi_1-A\cos(\psi_1+\theta_1-60^\circ)} \tag{28}$$

where

$$A=\frac{\sin\psi_1}{\sqrt[3]{T(0)}\sin(\psi_1+\theta_1)} \tag{29}$$

Equation 23 may be used to determine θ_2 with θ_1 and θ_3 known.

By consideration of equation 27 plus the condition 23, in view of the requirement that $\theta_3>\theta_2>\theta_1$, which follows from Fig. 7(D), it is found that θ_1 cannot exceed 60 degrees and that β_1 attains its maximum value for a given $T(0)$ as θ_1 approaches this figure. To provide an indication of the relative values involved in this stabilization method, θ_1 was selected as 55 degrees. θ_2 and θ_3 were found to be 115.0 degrees and 110.0 degrees, respectively, which then gave values of $m=3.28$, $\rho_1=0.397\rho_0$, $\alpha_1=1.384\ \rho_0$, and $\beta_1=2.70\rho_0$. The values of R_1 and C_1 necessary to produce this performance can be found using equations 19 and 20 when R and ρ_0 are known.

Local Feedback—Original Asymptote

The common emitter transistor amplifier circuit shown in Fig. 8(A) which has been discussed by Abraham[12] and Blecher[13,14] provides an example of the way in which local feedback can be used to achieve pole-zero patterns suitable for stabilization. The straight line approximation to the current gain I_2/I_1 is shown in Fig. 8(B). If R_2 is much greater than the base resistance r_b' of the transistor, then the operation with feedback is along the original asymptote beyond the frequency f_1'. Without feedback, the amplifier has the pole pattern shown in Fig. 8(C); with the elements R_3-C_3 added, the pattern of Fig. 8(D) results. Here again a constraint on pole-zero locations is noted. By choice of the product R_3C_3, the real zero ($\rho_3=1/R_3C_3$) can be located arbitrarily. The addition of the network R_3C_3 adds a pole at $-\rho_3$ and shifts the locations of the poles at $-\rho_0$ and $-\rho_1$ to $-\rho_0'$ and $-\rho_1'$. The values of ρ_0', ρ_1', and ρ_2 depend upon the choice of R_3 and C_3. Blecher[14] has given the following approximate values for ρ_0 and ρ_1, (see reference 12 for details of the approximations):

$$\rho_0=\frac{1-a_0+\delta}{\dfrac{1+a_0m+\delta}{\omega_a}+\dfrac{1}{\omega_c}} \tag{30}$$

$$\rho_1=\omega_a+\omega_c(1+a_0m+\delta) \tag{31}$$

If ρ_3 is selected to be much less than ρ_0' (as given next) then the approximate locations of the poles with local feedback are given by

$$\rho_0'=\frac{1-a_0+\gamma}{\dfrac{1+a_0m+\gamma}{\omega_a}+\dfrac{1}{\omega_c}} \tag{32}$$

$$\rho_1'=\omega_a+\omega_c(1+a_0m+\gamma) \tag{33}$$

$$\rho_3=\frac{1-a_0+\delta}{1-a_0+\gamma}\,\frac{1}{C_3R_3} \tag{34}$$

In all of these equations

$$\delta=\frac{R_1+r_e}{r_e};\quad \gamma=\frac{R_1+r_e}{R_3+r_b'};\quad \omega_c=\frac{1}{(R_1+r_e)C_c} \tag{35}$$

and C_c, r_b', r_e, r_c, and a_0 are the usual transistor parameters. m is the phase shift in radians by which the phase of a exceeds the phase shift of an ordinary R-C network at the frequency ω_a where the amplitude of a is 0.707 a_0. A measurement technique for the determination of m has been discussed by Thomas and Moll.[15]

From the similarity of the straight line approximations of Fig. 7(B) and Fig. 8(B) and the corresponding pole-zero patterns, it is clear that one may proceed to stabilize a transistor feedback amplifier using a transient response criterion in essentially the same fashion as in the previous example. (Pederson and Ghausi[16] have used the root locus method to design a transistor feedback amplifier having a prescribed closed-loop amplitude-frequency response characteristic.) There are two complications which may arise, however. It is seen from equations 32, 33, and 35 that ρ_0' and ρ_1' depend on R_3 through the factor γ and therefore, the locations of ρ_0' and ρ_1' relative to ρ_0 and ρ_1 are not as simply expressed as in the previous example. In addition, if the effect of the excess phase shift which is represented approximately by a factor e^{-jm/ω_a} in the transfer function is large enough, the root locus should be modified as considered by Chu.[17] It should be noted that both of these possible complications arise because of the nature of the device and the particular circuit employed; the factors arise similarly in a

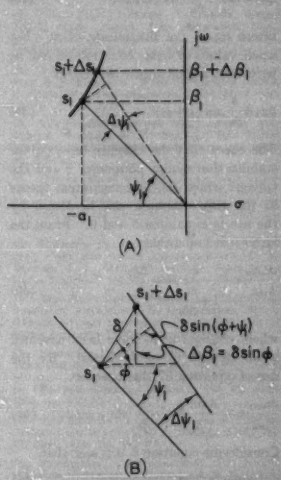


Fig. 9. Effect of parameter variations

A—Dominant pole at $s_1=-\alpha_1+j\beta_1$ and shift in pole location (Δs_1) as a result of parameter change

B—An expanded portion of (A) illustrating the geometrical relations associated with the change in dominant pole location

stabilization design procedure based on frequency response as well as in one based on transient response.

Effect of Parameter Variations

Changes in the active as well as passive elements comprising a feedback amplifier can cause a shift in the location of the open loop, and hence closed loop, pole, and zero locations. It is of interest to consider briefly what effect these variations have on the transient response of the amplifier. Fig. 9(A) shows the variation of one dominant pole position from s_1 to $s_1+\Delta s_1$. Such a change might be caused by a change in magnitude of mid-band return ratio, for example, in which event the broad segment shown would be part of a root locus. In general, a change in dominant pole location causes both ψ_1 and β_1 to change. From Fig. 9(B) which is an expanded portion of Fig. 9(A), it is seen that if the change in dominant pole location is expressed as

$$\Delta s_1 = \delta e^{j\varphi} \tag{36}$$

then

$$\Delta\beta_1 = \delta \sin \varphi \tag{37}$$

and

$$\Delta\psi_1 \doteq \tan^{-1} \frac{\delta}{\beta_{01}} \sin(\varphi+\psi_1) \tag{38}$$

where β_{01} is the magnitude of s_1. For small values of δ/β_{01}, equation 38 can be written to a good approximation as

$$\Delta\psi_1 \doteq \frac{\delta}{\beta_{01}} \sin(\varphi+\psi_1) \tag{39}$$

The effect that the changes have on the stabilization design parameter η and the time at which the first maximum occurs in the time response can be found by the use of equations 5 and 6. From the latter it is found that

$$\frac{\Delta t_{max}}{t_{max}} = -\frac{1}{1+\dfrac{\beta_1}{\Delta\beta_1}} \doteq -\frac{\Delta\beta_1}{\beta_1} \tag{40}$$

where changes in λ_1 have been assumed small enough to be neglected. By the use of equation 37 this becomes

$$\frac{\Delta t_{max}}{t_{max}} \doteq -\frac{\delta \sin\varphi}{\beta_1} \doteq -\frac{\delta}{\beta_{01}} \sin\varphi \csc\psi_1 \tag{41}$$

Considering equation 5, it is seen that

$$\eta+\Delta\eta = e^{-\sigma \cot(\psi_1+\Delta\psi_1)} \tag{42}$$

Since $\cot(\psi_1+\Delta\psi_1)$ can be expressed in the Taylor's series.

$$\cot(\psi_1+\Delta\psi_1) = \cot\psi_1 - \Delta\psi_1 \csc^2\psi_1 + (\Delta\psi_1)^2 \csc^2\psi_1 \cot\psi_1 + \dots \tag{43}$$

the fractional change in η can be written approximately as

$$\frac{\Delta\eta}{\eta} = e^{\sigma\Delta\psi_1 \csc^2\psi_1} - 1 \tag{44}$$

where only the linear term in $\Delta\psi_1$ in the series has been used. Use of equation 39 allows the change in η to be written as a function of the change in s_1, viz.

$$\frac{\Delta\eta}{\eta} = -1 + e^{\sigma\delta/\beta_{01} \csc^2\psi_1 \sin(\varphi+\psi_1)} \tag{45}$$

The changes in t_{max} and η may be related to parameter changes by application of results reported in three recent papers. Huang[18] has studied the change in closed-loop pole locations as a result of variations in open loop pole and zero locations or gain constant. Truxal and Horowitz[19] considered the sensitivity of a pole of a return difference to any parameter in the network under consideration, and their work was extended by Kuo[20] to the sensitivity of a root of a polynomial with respect to any network parameter.

Consider an amplifier characterized by equation 8. To study the effect of the variation in the magnitude of the mid-band return ratio $T(0)$ on Δs_1 and hence η and t_{max}, use may be made of equation 10 of Huang's paper. In terms of the notation here, it is found that

$$\Delta s_1 = -\frac{\Delta T(0)}{T(0)} k_{-1} \tag{46}$$

where k_{-1} is the residue of the closed loop gain function at the pole s_1. Expression of this result in the form of equation 36 allows evaluation of equations 41 and 44 by substitution of δ and φ. Similar calculations may be made to study the effect on the transient response of other parameter variations but the detailed computation is beyond the scope of this paper.

Conclusions

By the combination of root-locus methods and the concept of the dominance of the step-function response, it is possible to formulate a stabilization design criterion for feedback amplifiers based on direct control of the transient response. The design criterion suggested, the damping or rate of attenuation of the dominant time term in the step-function response, is a simple function of the pole angle of the dominant pair of complex poles and is also related to the maximum transient overshoot. Illustrative examples indicate how one may proceed directly from the statement of desired performance in terms of the criterion to element values in the stabilization networks.

The problem of optimizing the design of a feedback amplifier for a prescribed transient response differs from the corresponding servomechanism problem because of the constraints imposed by the characteristics of the active devices employed and their associated coupling networks. Further investigation in which design methods already known for achieving prescribed transient response with servomechanisms are coupled with the results of studies of constraints imposed on amplifier interstage pole-zero configurations by the addition of stabilization networks should be profitable in increasing the applicability of the procedure suggested. A stabilization philosophy for feedback amplifiers based directly on control of transient response can, as it is developed further, provide the circuit designer with far greater insight into the effect of design parameters than criteria based solely on gain and phase margins.

Appendix

Derivation of Equations 14 and 15

In Fig. 6(D), let γ be the angle that the directed distance from $-a\rho_0$ to $-\alpha_1+j\beta_1$ makes with the positive real axis. In order to satisfy the condition that the angle of $T'(s)$, equation 13, is equal to -180 degrees, it is necessary that

$$2\theta+\gamma = 180 \text{ degrees}$$

or

$$\gamma = 180-2\theta$$

Thus, the supplement of γ has the value 2θ which is shown in Fig. 6(D).

Because of the condition $|T'(s)| = 1$ at $s = -\alpha_1+j\beta_1$, equation 13 may be reduced to the form

$$1 = \frac{T(0)}{\left[\dfrac{|s+\rho_0|}{\rho_0}\right]^3 \dfrac{|s+a\rho_0|}{a\rho_0}}$$

The factors $|s+\rho_0|$ and $|s+a\rho_0|$ are recognized as the magnitudes of the directed distances from $-\rho_0$ to $-\alpha_1+j\beta_1$ and from $-a\rho_0$ to $-\alpha_1+j\beta_1$, respectively. By application of the law of sines to the triangles in the figure, it is found that

$$\frac{|s+\rho_0|}{\rho_0} = \frac{\sin\psi_1}{\sin[180-(\theta+\psi_1)]} = \frac{\sin\psi_1}{\sin(\theta+\psi_1)}$$

$$\frac{|s+a\rho_0|}{a\rho_0} = \frac{\sin\psi_1}{\sin[180-(\gamma+\psi_1)]} = \frac{\sin\psi_1}{\sin(2\theta-\psi_1)}$$

Substitution of these results in the previous expression containing $T(0)$ yields

$$\frac{\sin^3\psi_1}{\sin^3(\theta+\psi_1)} \frac{\sin\psi_1}{\sin(2\theta-\psi_1)} = T(0)$$

from which equation 14 follows directly.

To obtain equation 15 for a, the condition above derived from equation 13 may first be written as

$$a = \frac{1}{T(0)} \left[\frac{|s+\rho_0|}{\rho_0}\right]^3 \frac{|s+a\rho_0|}{\rho_0}$$

$$= \frac{1}{T(0)} \frac{\sin^3\psi_1}{\sin^3(\theta+\psi_1)} \frac{|s+a\rho_0|}{\rho_0}$$

If ρ_{01} denotes the magnitude of the directed distance from $-\rho_0$ to $-\alpha_1+j\beta_1$, then

$$\frac{|s+a\rho_0|}{\rho_0}=\frac{|s+a\rho_0|}{\rho_{01}}\,\frac{\rho_{01}}{\rho_0}$$

Application of the law of sines yields the equality

$$\frac{|s+a\rho_0|}{\rho_0}=\frac{\sin\theta}{\sin 2\theta}\,\frac{\sin\psi_1}{\sin(\theta+\psi_1)}=\frac{\sin\psi_1}{2\cos\theta\sin(\theta+\psi_1)}$$

Substitution of the expression for this ratio in terms of θ and ψ_1 in the equation for a results in equation 15, viz.

$$a=\frac{1}{2T(0)\cos\theta}\,\frac{\sin^2\psi_1}{\sin^2(\theta+\psi_1)}$$

Derivation of Equations 22 and 24 through 28

The magnitude condition $|T'(s)|=1$ applied to equation 21 results in the following equality:

$$|s+m\rho_0|^2|s+\rho_1/m|^2=\rho_0^4T(0)|s+\rho_1|^2$$

An equivalent form is

$$\frac{|s+m\rho_0|}{m\rho_0}\,\frac{|s+\rho_1/m|}{\rho_1/m}=\sqrt[4]{T(0)}\,\frac{|s+\rho_1|}{\rho_1}$$

The use of the law of sines in the appropriate triangles of Fig. 7(D) allows the ratios of magnitudes to be expressed in terms of the angles θ_1, θ_2, θ_3, and ψ_1 as follows:

$$\frac{|s+m\rho_0|}{m\rho_0}=\frac{\sin\psi_1}{\sin[180-(\psi_1+\theta_1)]}=\frac{\sin\psi_1}{\sin(\psi_1+\theta_1)}$$

$$\frac{|s+\rho/m|}{\rho_1/m}=\frac{\sin\psi_1}{\sin(\psi_1+\theta_2)};$$

$$\frac{|s+\rho_1|}{\rho_1}=\frac{\sin\psi_1}{\sin(\psi_1+\theta_3)}$$

Substitution of these results in the equality above changes its form to

$$\frac{\sin\psi_1}{\sin(\psi_1+\theta_1)}\,\frac{\sin\psi_1}{\sin(\psi_1+\theta_1)}=\sqrt[4]{T(0)}\,\frac{\sin\psi_1}{\sin(\psi_1+\theta_3)}$$

which may be converted directly to equation 22.

By examination of Fig. 7(D), it is seen that the law of sines may be used to express the ratio $|s+\rho_1|/|s+\rho_1/m|$ in terms of θ_3 and θ_2. The equality is

$$\frac{|s+\rho_1|}{|s+\rho_1/m|}=\frac{\sin\theta_2}{\sin\theta_3}$$

One may then start with the equality derived above from the condition $|T'(s)|=1$ in the form

$$m\frac{|s+m\rho_0|}{m\rho_0}=\sqrt[4]{T(0)}\,\frac{|s+\rho_1|}{|s+\rho_1/m|}$$

and substitute values for the ratios in terms of the various angles in the figure to obtain

$$m\frac{\sin\psi_1}{\sin(\psi_1+\theta_1)}=\sqrt[4]{T(0)}\,\frac{\sin\theta_2}{\sin\theta_3}$$

This gives equation 24 for m, namely

$$m=\sqrt[4]{T(0)}\,\frac{\sin(\theta_1+\psi_1)}{\sin\psi_1}\,\frac{\sin\theta_2}{\sin\theta_3}$$

Suppose β_{01} designates the distance from the origin to $-\alpha_1+j\beta_1$. Then proceeding as previously,

$$\frac{\beta_{01}}{\rho_1}=\frac{\sin\theta_2}{\sin[180-(\psi_1+\theta_2)]}=\frac{\sin\theta_2}{\sin(\psi_1+\theta_2)}$$

$$\frac{\beta_{01}}{m\rho_0}=\frac{\sin\theta_1}{\sin(\psi_1+\theta_1)}$$

whence

$$\rho_1=m\rho_0\,\frac{\sin\theta_1}{\sin(\psi_1+\theta_2)}\,\frac{\sin(\psi_1+\theta_2)}{\sin\theta_3}$$

which is equation 25.

The quantities α_1 and β_1 can be expressed simply in terms of β_{01} and ψ_1 as

$$\alpha_1=\beta_{01}\cos\psi_1$$

$$\beta_1=\beta_{01}\sin\psi_1$$

Substitution of the value for β_{01} in terms of $m\rho_0$ gives the forms

$$\alpha_1=m\rho_0\,\frac{\sin\theta_1\cos\psi_1}{\sin(\psi_1+\theta_1)}$$

$$\beta_1=m\rho_0\,\frac{\sin\theta_1\sin\psi_1}{\sin(\psi_1+\theta_1)}=\rho_0\sqrt[4]{T(0)}\,\frac{\sin\theta_1\sin\theta_2}{\sin\theta_3}$$

These are equations 26 and 27; equation 24 for m has been used in passing to the second form of the expression for β_1.

Equation 28 is derived by starting with equations 22 and 23, with n taken equal to zero and the plus sign chosen in the latter, that is,

$$\theta_3=\theta_1+\theta_2-60$$

Equation 22 can now be written as

$$\frac{\sin(\psi_1+\theta_3)}{\sin(\psi_1+\theta_1+\theta_2-60)}=\frac{\sin\psi_1}{\sqrt[4]{T(0)}\sin(\psi_1+\theta_1)}=A$$

Clearly A is a known quantity since ψ and $T(0)$ are assigned, and it is assumed that a value has been chosen for θ_1. The terms $\sin(\psi_1+\theta_3)$ and $\sin(\psi_1+\theta_1+\theta_2-60)$ may be expanded to

$$\sin(\psi_1+\theta_3)=\sin\psi_1\cos\theta_1+\cos\psi_1\sin\theta_3$$

and

$$\sin(\psi_1+\theta_1+\theta_2-60)=\sin(\psi_1+\theta_1-60)\cos\theta_2+\cos(\psi_1+\theta_1-60)\sin\theta_2$$

The previous equation,

$$\sin(\psi_1+\theta_3)=A\sin(\psi_1+\theta_1+\theta_2-60)$$

written in terms of the expansions just provided becomes

$$[\cos\psi_1-A\cos(\psi_1+\theta_1-60)]\sin\theta_2=-[\sin\psi_1-A\sin(\psi_1+\theta_1-60)]\cos\theta_2$$

This may be rewritten in the form of equation 28.

References

1. Network Analysis and Feedback Amplifier Design (book), H. W. Bode. D. Van Nostrand Company, Princeton, N. J., 1945.

2. Some Considerations in the Design of Negative-Feedback Amplifiers, W. T. Duerdoth. *Proceedings*, Institution of Electrical Engineers, London, England, pt. III, vol. 97, 1950, pp. 133–58.

3. A Simple Connection Between Closed-Loop Transient Response and Open-Loop Frequency Response, J. C. West, J. Potts. *Ibid.*, pt. II, vol. 100, 1953, pp. 201–12.

4. Servomechanism Transient Performance from Decibel-Log Frequency Plots, H. Harris, Jr., M. J. Kerby, E. F. VonArx. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1452–59.

5. Control System Dynamics (book), W. R. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

6. The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems, J. H. Mulligan, Jr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, 1949, pp. 516–29.

7. Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed Loop Function, M. R. Aaron. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1439–46.

8. Feedback System Synthesis by the Inverse Root-Locus Method, J. A. Aseltine. *Convention Record*, Institute of Radio Engineers, vol. 4, pt. 2, 1956, pp. 13–17.

9. Synthesis of Feedback Control System by Phase-Angle Loci, Y. Chu. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 71, Nov. 1952, pp. 330–39.

10. Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location, D. W. Russell, C. H. Weaver. *Ibid.*, Jan., pp. 95–104.

11. Integrated s-Plane Synthesis Using 2-Way Root Locus, John Zaborszky. *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, 1956 (Jan. 1957 section), pp. 797–801.

12. A Wide-Band Transistor Feedback Amplifier, R. P. Abraham. *Wescon Convention Record*, Institute of Radio Engineers, 1957, pt. 2, pp. 10–19.

13. Transistor Circuits for Analog and Digital Systems, F. H. Blecher. *Bell System Technical Journal*, New York, N. Y., vol. 35, 1956, pp. 295–332.

14. Design Principles for Single Loop Transistor Feedback Amplifiers, F. H. Blecher. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. CT-4, 1957, pp. 145–56.

15. Junction Transistor Short-Circuit Current Gain and Phase Determination, D. E. Thomas, J. L. Moll. *Proceedings*, Institute of Radio Engineers, vol. 46, 1958, pp. 1177–84.

16. The Root Locus Design of Transistor Feedback Amplifiers, D. O. Pederson, M. S. Ghausi. *Wescon Convention Record*, Institute of Radio Engineers, 1958, pt. 2, pp. 87–93.

17. Feedback Control Systems With Dead-Time Lag or Distributed Lag by Root-Locus Method, Yaohan Chu. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 71, Nov. 1952, pp. 391–96.

18. The Sensitivity of the Poles of Linear, Closed-Loop Systems, R. Y. Huang. *Ibid.*, vol. 77, Sept. 1958, pp. 182–87.

19. Sensitivity Considerations in Active Network Synthesis, J. G. Truxal, I. M. Horowitz. *Proceedings*, Second Midwest Symposium on Circuit Theory, Michigan State University, East Lansing, Mich., Dec. 1956.

20. Pole-Zero Sensitivity in Network Functions, F. F. Kuo. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. CT-5, 1958, pp. 372–73.

Proposed Recommended Practices for Medium-Voltage Motor Controllers for Rubber and Plastics Industries

WILLIAM S. WATKINS
MEMBER AIEE

THE RUBBER and Plastics Subcommittee of the General Industry Applications Committee is presenting for discussion the following proposed Recommended Practices, in the hope that comments from the wide circulation so provided will be helpful in attaining maximum usefulness from the specifications. Membership of the Working Group assigned to this project consists of engineers from both large and small rubber companies, electrical manufacturers, machinery builders, and consulting engineering firms. The requirements listed are a composite of the desires of the persons who apply controls to the large rubber and plastics processing machines, influenced to an appropriate degree by the manufacturers. Discussions resulting from this presentation are necessary to assure the widest possible acceptance of the proposal.

It has been a basic consideration in the deliberations of the Working Group that new standards should depart from existing standards only where absolutely necessary, and it is hoped that these proposed Recommended Practices might possibly serve as a basis for standards acceptable to the entire processing industry, with deviations only where peculiar conditions make some industries differ greatly from others. Comments will be welcomed on this general aspect as well as on the specific suitability of the recommendations.

The following is to be proposed to the AIEE Standards Committee as a "Recommended Practice" as described in Division III, Section I of the AIEE Standards Manual. It is not intended to be an AIEE Standard, and the recommendations are not mandatory or restrictive.

Foreword

The advantages of standardization in this field have been considered at sessions of the Rubber and Plastics Subcommittee since 1948. At the 1956 conference in Akron, Ohio, it was voted that a Working Group should be appointed to draft standards for the industries, and that this group should direct its efforts first toward the problem of the medium-voltage motor controller. Progress reports were made at the 1957 and 1958 conferences, and in 1958 the subcommittee voted unanimously to support the Recommended Practices proposed by the Working Group.

Throughout the period required for the development of the Recommended Practices it has been the policy of the Working Group to invite participation by all interested engineers, and comments have been received from many who were not actually members of the group. A liaison committee representing the General Engineering Committee of the National Electrical Manufacturers Association worked closely with the AIEE group, and is proposing the incorporation of parts of these Recommended Practices as National Electrical Manufacturers Association (NEMA) standards. The Electrical Engineering Committee of the Technical Association of the Pulp and Paper Industry (TAPPI) has followed the progress of this project, and has proposed for its industry a set of Recommended Practices almost identical to these.

It has been the intent of the Working Group to make this publication as useful as possible at the earliest possible date. It is assumed that the Rubber and Plastics Subcommittee will make certain that the very latest and best thinking is incorporated into these Recommended Practices from time to time, in order to encourage rather than hinder progress.

Section 1. General

1.1 PURPOSE The purpose of these Recommended Practices is to define and recommend specifications for medium-voltage motor controllers for application in the rubber and plastics industries, in order to promote:

(a). Safety to personnel and equipment, including the reduction of fire and accident hazards.

(b). Maximum production with minimum unscheduled shutdowns.

(c). Reduced maintenance and increased life of equipment.

(d). Clarification of needs and conditions, to reduce special engineering and chance of error in specifications.

(e). Over-all economy.

The contents of this publication are intended to bring to the attention of engineers, management, and equipment suppliers the important features necessary to meet the specific requirements of the rubber and plastics industries.

1.2 SCOPE The provisions of these Recommended Practices apply to the construction and application of controllers for motors in the medium-voltage classification (601 to 5,000 volts.)

Although many of the specifications herein have been influenced by the conditions peculiar to the processing areas of these industries, it is probable that they may be applied to all purchases of controllers for a given company in order to promote uniformity. They are intended as minimum requirements, and may be supplemented where special needs exist.

1.3 USE OF RECOMMENDED PRACTICES It is urged that users and manufacturers co-operate and mutually agree to follow these recommended practices to insure the use of proven apparatus and methods for economical and safe machine operation.

1.4 APPLICABLE STANDARDS It is assumed that apparatus will comply with the National Electrical Code and NEMA standards insofar as they are applicable.

Section 2. Enclosure

2.1 DESCRIPTION Enclosure shall be general-purpose NEMA 1. High-voltage compartment shall be segregated. Material which will support combustion shall be kept to a minimum.

2.2 DIMENSIONS Basic enclosure shall have a height of 90 inches exclusive of mounting sills. Preferred depth is 30 inches; maximum depth is 32 inches.

2.3 POWER AND CONTROL LEADS Space for conduit or armored cable entrance shall be provided at top and bottom.

2.4 LOCATION OF RESISTORS Resistors which require ventilation shall be mounted

---

Paper 59-1112, recommended by the AIEE General Industry Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted June 15, 1959; made available for printing July 15, 1959.

WILLIAM S. WATKINS is with The Ohio Rubber Company, a division of Eagle-Picher Company, Willoughby, Ohio.

in a ventilated portion of the enclosure, or externally mounted.

Section 3. Power Busses

3.1 Location Busses, if specified, shall be on top of the control enclosure at the front. All manufacturers shall locate the busses in the same position so controllers of different manufacturers can be mounted adjacent to each other and the busses can be easily connected between controllers.

3.2 Enclosure Bus compartment shall be isolated from the controller interior.

3.3 Splices Bus-bar connections shall be silver plated.

Section 4. Isolating Device

4.1 Description The isolating device shall be a 3-pole gang-operated isolating switch. With compartment door open, the blades shall be visible when open, or else disconnecting-type fuses shall be furnished. Operating handle of this switch shall be outside of compartment.

4.2 Padlocking Provision shall be included to accommodate three padlocks for locking the switch in the open position.

Section 5. Interrupting Rating

5.1 Interrupting Rating The interrupting rating shall be at least 150,000 kva at 2,300 volts and 250,000 kva at 4,600 volts.

5.2 Description Power fuses, if used, shall be of the current-limiting type.

Section 6. Contactors

6.1 Description Contactors shall be of the air-break type and not mechanically latched in the closed position. They shall be either removable or readily accessible.

Section 7. Accessibility

7.1 Accessible from Front All functioning parts of the controller shall be readily accessible for maintenance or replacement from front of the controller.

7.2 Low-Voltage Panel Low-voltage panel shall either be front wired, hinged, or otherwise accessible for maintenance and inspection of all wiring.

7.3 Terminal Boards All terminals for outgoing low-voltage conductors to be at least 12 inches from floor and readily accessible for connection and identification. Sufficient space shall be provided for routing of customer's control wiring, segregated from high-voltage leads, to such terminals boards. On contactors, terminal boards for coil leads and control wiring shall be readily accessible.

Section 8. Interlock Requirements

8.1 Isolating Switch and Compartment Door Isolating switch must be open before any high-voltage compartment door can be opened and cannot be closed before all such compartment doors have been closed.

8.2 Isolating Switch and Contactors It shall not be possible to open or close the isolating switch while any of the line contacts are closed, or closing.

8.3 Contactors The *F*, *R*, and *DB* contactors shall be mechanically interlocked. The *DB* contactor shall be preferably interlocked in most direct manner with the *F* contactor.

8.4 Provision for Testing Provision shall be made so that controller can be operated for testing, but only with main isolating switch open. If a "Test-Normal" switch is specified, it shall be arranged to nullify control stations remote from the controller when testing and to nullify test push buttons at the controller when in "Normal position, except that all "Stop" push buttons, whether for "Normal" or "Test" operation, shall be connected in series, so as to be effective at all times. The Test switch shall also transfer the control circuit from the normal source to a test power source to be furnished by user. Test push buttons shall be located inside the controller.

8.5 Types of Interlocks Mechanical-type interlocks are preferred over key-type interlocks. Some means shall be provided for gaining entry to the enclosure in case a contactor should weld closed.

Section 9. Protective Devices and Control Circuits

9.1 Protective Devices Control shall include control devices such as undervoltage relay, overload relays, field application where required, etc., which are recognized as necessary for adequate protection to personnel and equipment.

9.2 Excitation Failure A field-loss relay, power-factor relay, or other effective means shall be provided to insure the shutdown of synchronous motor in the event of excitation failure. This device shall operate on loss of d-c excitation voltage or current, and shall prevent restarting until it has been manually reset. The resetting means shall be clearly marked as to function.

9.3 Number of External Wires Control circuit shall be designed so that a minimum number of wires to external devices are required.

9.4 A-C Control Circuit. When a-c control is required, it shall be supplied by a control transformer, 220 volts secondary, on each controller. It shall be supplied from the load side of the main disconnect through its own current-limiting fuses. A-c control shall also be provided with main disconnecting means and overcurrent protection on secondary side. Provision shall be made for grounding one side of secondary of control transformer and where ground bus is specified this connection shall be made to the ground bus.

9.5 D-C Control Circuit When d-c control is required it shall be supplied through a disconnecting switch and fault current protective means furnished as part of the controller. When d-c excitation is required the controller shall include means for disconnecting, short-circuit protection, and adequate field discharge protection. Adequate field discharge protection is understood to include a discharge means permanently connected across the load side of the circuit protective device, in addition to any other resistors which may be connected to the field circuit on occasion by discharge clips or normally closed contacts.

9.6 Emergency Stop If emergency braking is required, emergency stop switches shall break both sides of the control circuit (American Standard safety code for rubber mills and calenders *A2.1.6*). The time required between safety switch opening and braking contactor closing shall be the minimum possible.

All "Stop" push buttons, as well as all "Emergency Stop" switches, shall cause a braked stop. Means shall be provided to ensure completion of the normal braking cycle in all cases before restarting.

9.7 Conductors All control conductors shall be stranded.

9.8 Undervoltage Protection Instantaneous undervoltage protection shall be furnished unless time-delay is specified.

9.9 Fusing of Potential Transformers Potential transformers, if fur-

nished, shall be fused primary and secondary. Primary fuses shall be of the current-limiting type.

Section 10. Name Plates

10.1 DESCRIPTION Permanent and legible name plates shall be provided for all devices, with designations per NEMA Standards *ICI-3.01*.

10.2 POWER FUSE DESIGNATION Nameplate showing catalog number of power circuit fuses used shall be mounted on inside of fuse compartment door.

Section 11. Wiring Diagram

11.1 LOCATION OF CONTACTS IN CONTROL CIRCUIT On schematic diagram, all contacts shall be on left-hand side of coils, except overload contacts and emergency stop switches which break both sides of line.

11.2 GENERAL REQUIREMENTS Wiring diagram shall include schematic and panel diagram. Each line of schematic diagram shall be numbered on left-hand side, and location of each control circuit contact on a device designated on the right-hand side adjacent to the coil. The line numbers of normally closed contacts should be underlined.

Discussion

E. H. Peters (Westinghouse Electric Corporation, Chicago, Ill.) and **W. F. Gardner** (Westinghouse Electric Corporation, Akron, Ohio): In our opinion the committee accomplished what they set out to do, that is, to provide a specification that would give a standard acceptable to both users and industry. We believe most users were able to have included in the specifications the features which they desired. The features where users did not agree usually represented the result of different company practices and were recognized as such. In those cases, each user will specify the modifications he requires. Most suppliers, we believe, feel the specification is not restrictive to prevent the use of individual design initiative.

There seem to be two items however in which compromise solutions were accepted which were not entirely satisfactory to suppliers or users. These were:

1. The present bus location allows only 6 inches of space at the back of the cabinet to bring out motor leads. This is certainly the minimum acceptable. Since it appears that bus location will ultimately become a NEMA standard perhaps a better solution of this problem can be found.

2. Various methods of protection have been offered by suppliers to shut down the motor in the event of loss of field. Nothing, however, was available to protect the machine if the field circuit was open and the operator attempted to start. It was agreed that this was the operator's usual reaction if he could not find what shut the machine down. A compromise was reached whereby the field-loss relay would operate a manual reset device thereby warning the operator of the field loss.

Since the writing of these specifications a circuit has been developed which actually samples the continuity of motor field and starting resistor circuit before the main contactors can be picked up. This eliminates the necessity of having a manually reset field-loss relay and also protects the machine from incorrect wiring at the time of start up and the loss of fields due to raised brushes. As this circuit receives further field testing it should be considered as an improved feature to be included in the specifications.

William S. Watkins: Oral comments and the written discussion received seem to verify that the proposal is generally acceptable, with the reservation that a few points should be subjected to a little more scrutiny. Questions were raised concerning the grounding of one side of the control transformer, the location of the supply busway, and the provision for protection against lack of field continuity. These are certainly very penetrating questions, and will be studied by the Working Group before the final draft is submitted to the Standards Committee. We shall be grateful for any additional comments which may be submitted, on these and other items, to assist the group in making its decisions.

Use of a Mathematical Error Criterion in the Design of Adaptive Control Systems

C. W. MERRIAM, III
ASSOCIATE MEMBER AIEE

Synopsis: Mathematical error criteria are examined in terms of the goals of an adaptive control system. For dynamic processes which contain pure-time delay, a modification of these error criteria is postulated on the basis of the time-domain concept of control developed here.

Also, the mathematical methods used in optimization are examined in terms of the feedback concept. In particular, the optimum configuration of the control system is discussed and illustrated for dynamic processes with and without pure-time delay.

THE GENERAL topic of adaptive control systems presently is of great interest to engineers who traditionally have employed the concepts and methods of feedback control. In this regard, much of the impetus stems from the observation that, even though the human lacks many attributes required of purely physical systems such as speed of response, the human does possess desirable characteristics which are as yet unattainable in a purely physical form. Many of these desirable human characteristics arise from subjective judgment and the ability to utilize *a priori* information in situations identifiable by the human as having been encountered previously.

In an effort to emulate the subjective characteristics sometimes attributed to the human, researchers have postulated classes of systems such as self-adjusting and self-organizing which broadly could be classified as adaptive. Unfortunately, very little is understood regarding human behavior. Hence, systems designed to approximate human characteristics are restricted from a philosophical standpoint by the validity of the model of the human adopted, even though the response characteristics of the resulting systems may be very acceptable from some objective viewpoint.

In order to circumvent such philosophical arguments, this paper is predicated on the use of a mathematical index of performance for the synthesis of a class of adaptive control systems. The author regards this class of adaptive control systems as particularly useful in practice because adaptation is accomplished in an optimum fashion merely by gain changes. Furthermore these gains can be computed readily so that real-time computa-

Paper 59-1158, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted April 8, 1959; made available for printing August 13, 1959.

C. W. MERRIAM, III, is with Massachusetts Institute of Technology, Cambridge, Mass.

This work was supported in part by the International Business Machines Company (IBM) under IBM Purchase Order No. L-R19014, and under the Massachusetts Institute of Technology DSR Project No. 7793.

tions are feasible and could be employed to make these adaptive controllers applicable in a variety of situations

Selection of an Error Criterion

Generally speaking, performance indices may be put into two categories. The first category is defined here as the shaping of transient response characteristic without regard to the nature of the particular response desired of the system. For instance, a shaping criterion may be the specification that the closed-loop system should approximate a linear, second-order system with a given rise time, setting time, or possibly a given overshoot. Unfortunately, many systems contain time varying and nonlinear components which are not readily amenable to these simple conceptual interpretations of the second-order system. The second category is defined here as the optimum approximation of desired response signals where errors in the approximation are weighted according to an appropriate performance criterion, called an error criterion when used in this sense. Here, then, an optimum occurs when the weighted error is minimized, thereby specifying the transient response characteristics of the resulting system in addition. For this paper, only performance indices of the second category are discussed in order to avoid the inevitable limitation of attempting to specify appropriate criteria that are based on linear concepts for the design of adaptive systems which are by definition nonlinear and time varying.

Having adopted the method of designing an adaptive system by the minimization of an error criterion, the specific class of design problems which reflect the goals of an adaptive control system must be defined. Here an adaptive control system is intended to be a feedback system which is altered continuously in order to maintain optimum system response. Controller alterations are necessary when changes occur in the estimated characteristics of the controlled dynamic process and of the response desired of the system. Also, alterations are necessary when the error criterion is changed. It should be mentioned here that the scope of this paper is restricted, for the sake of brevity, by neglecting the errors which may occur in the estimation of both the characteristics of the dynamic process and of the desired response. Inclusion of such errors is a study in itself and suitable results have been obtained only when the errors occur additively in the control and response signals.

A natural starting point for the selection of an error criterion might be the use of a least-squares criterion which is traditional in approximation problems arising in many fields including numerical analysis. In particular, a valid error criterion is the integral

$$I=\int_{-\infty}^{\sigma}[Q(\sigma)-q(\sigma)]^2 d\sigma \qquad (1)$$

where $Q(\sigma)$ is the estimated desired system response signal, $q(\sigma)$ is the actual system response signal, and σ is a dummy time variable defined such that the value $\sigma=t$ is present or real time. The optimum system then would be defined as that system which minimizes the integral I. However, because practical components have limitations, a dynamic constraint must often be placed on the class of systems which are acceptable. A mathematically convenient way to insure the avoidance of saturating the control signal $m(\sigma)$ is the addition of a second term to the integral in equation 1, thereby, obtaining

$$I_c=\int_{-\infty}^{\sigma}[\lambda(\sigma)[Q(\sigma)-q(\sigma)]^2+m^2(\sigma)]d\sigma \qquad (2)$$

For the present, the assumption is made that the response signal $q(\sigma)$ and control signal $m(\sigma)$ are related unilaterally, as indicated in Fig. 1, by the linear ordinary differential equation

$$\sum_{u=0}^{U}a_u(\sigma)q^{(u)}(\sigma)=\sum_{u=0}^{U-1}b_u(\sigma)m^{(u)}(\sigma) \qquad (3)$$

The superscript (u) indicates the uth derivative of the function. Generally, the multiplier $\lambda(\sigma)$ in equation 2 is treated as an independent parameter such that $\lambda(\sigma)\geq 0$, and chosen so that $m^2(\sigma)\leq \mathfrak{M}^2(\sigma)$ where $\mathfrak{M}(\sigma)$ is the magnitude of the control signal where saturation occurs. This treatment of λ as a Lagrange multiplier is predicated, of course, on the proper engineering selection of the saturation level $\mathfrak{M}(\sigma)$.[1] Another interpretation of the criterion given in equation 2, however, is that the multiplier $\lambda(\sigma)$ can be determined directly from engineering considerations. A good example of this interpretation arises when the first term in equation 2 is representative of economic penalties arising from response errors and the second term in equation 2 is considered as the "cost" of control. For the purposes of this paper, no generality is sacrificed by assuming that the multiplier

Fig. 1. Block diagram of controlled process

$\lambda(\sigma)$ is determined directly from engineering considerations.

The error criteria presented thus far have no particular reference to adaptive control and are utilized conventionally, with the proper mathematical restrictions, in the design of linear time-invariant controls which of course are the antithesis of adaptive controls.

In order to postulate modifications to these criteria that are appropriate for application to adaptive controls, an analogy to the human faced with the problem of control is convenient. In words, the human attempts to select optimally the present value of the control signal $m(t)$. This selection is based on the existing conditions in the dynamic process, on the dynamic characteristics of the process, and on the response desired of the dynamic process in the future. Note that consideration of the future is necessary when the magnitude of the control system $m(\sigma)$ is constrained mathematically as in equation 2 because the existing conditions of the dynamic process cannot be changed instantaneously. Because of the conditions in the dynamic process, the dynamic characteristics of the process, and the desired response change with time, the optimum selection of the control signal changes with time. In this sense then, it is said that the human controller used in the analogy is adaptive.

The recognition that the existing conditions in the dynamic process are an integral part of adaptive control is of paramount importance to the selction of an appropriate error criterion. In particular, there is no apparent reason for weighting errors over the interval of past time $\sigma<t$ because, having occurred previously, no control can be exerted on them. If the assumption is made that the existing conditions in the dynamic process can be measured, as is desirous in a feedback control system, the errors that occur in the past can be neglected. Measurement of the existing conditions in the process is necessary in order to specify the effects of the past history of the process on the present and future response of the process. For processes which can be represented by an ordinary differential equation of finite order, elementary theory of such equations states that the effects of past history are specified completely by the present values of a finite number of state variables, usually called initial conditions. Now with the assumption that the state variables are measured, the error criterion used for adaptive control system design in this paper is

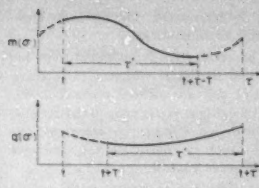


Fig. 2. Weighting intervals for dynamic process with pure-time delay of duration T

$$e=\int_{t}^{t+\tau}\{\lambda(\sigma)[Q(\sigma)-q(\sigma)]^2+m^2(\sigma)\}d\sigma \tag{4}$$

Note that the errors occuring in the interval of furture time $\sigma>t+\tau$ also have been neglected in equation 4. The omission of these future errors is necessary in many cases because adequate prediction or extrapolation of the desired response signal $Q(\sigma)$ may not be available. Also from a dynamic standpoint, no appreciable error is introduced into the selection of the optimum present value of the control signal $m(t)$ if the interval τ is chosen to be large in comparison to the time in which substantial transients can occur in the process. However, no particular restrictions are placed on the interval τ which is selected properly for the engineering problem at hand. For example, this interval may be chosen $\tau=\infty$ or $\tau=T_t-t$, where T_t is a constant as might be appropriate for control which terminates at time $t=T_t$.

Based on the heuristic assumption of the availability of a finite number of measurements which define the state of the dynamic process, the error criterion given in equation 4 may be appropriate for the adaptive control of a dynamic process which is described by an ordinary differential equation of finite order such as the linear one given in equation 3. However, many dynamic processes of interest involve pure-time delay due to distributed parameter effects where the measurement of the present values of a finite number of variables no longer represent completely the effects of past history on the present and future response of the process. A case in point is a dynamic process which is described by the differential equation

$$\sum_{u=0}^{U}a_u(\sigma)q^{(u)}(\sigma+T)=\sum_{u=0}^{U-1}b_u(\sigma)m^{(u)}(\sigma) \tag{5}$$

A comparison of equations 3 and 5 indicates that, for the same control signal $m(\sigma)$, the response signal of equation 5 is delayed by an interval of time T from that of equation 3. In addition to cases where equation 5 actually represents the dynamic process, an equation which contains pure-time delay but of low order often can be used as a good approximation to equation 3 when the order U is large. Such an approximation gives rise to a definite simplification in the optimum controller.

For dynamic processes with pure-time delay, appropriate changes in the error criterion given in equation 4 are facilitated by the sketches of the control and response signals given in Fig. 2. A pure-time delay T means that the segment of the control signal on the interval $t+\tau-T\leq\sigma\leq t+\tau$ is arbitrary in the sense that this segment does not affect the process response on the interval $t\leq\sigma\leq t+\tau$. Furthermore, the pure-time delay means that the control signal on the interval $t\leq\sigma\leq t+\tau$ does not affect the process response on the interval $t\leq\sigma\leq t+T$. Because the segment of the control signal on the interval $t+\tau-T\leq\sigma\leq t+\tau$ is arbitrary and because the segment of the response signal on the interval $t\leq\sigma\leq t+T$ is predetermined, a modified error criterion is postulated to be

$$e_T=\int_{t+\tau}^{t+\tau}\lambda(\sigma)[Q(\sigma)-q(\sigma)]^2d\sigma+\int_{t}^{t+\tau-T}m^2(\sigma)d\sigma \tag{6}$$

With a change in variables, this modified criterion can be written as

$$e_T=\int_{t}^{t+\tau'}\{\lambda(\sigma+T)[Q(\sigma+T)-q(\sigma+T)]^2+m^2(\sigma)\}d\sigma \tag{7}$$

where $\tau'=\tau-T$. This error criterion, then, weights only the portions of the signals that are meaningful, as determined from the time-domain concept developed here for adaptive control.

Determination of the Optimum Adaptive Controller

Even though an error criterion has been chosen which appears to be applicable to some adaptive control problems, useful results will be obtained only if the optimum controller can be specified explicitly. Consideration of the mathematical methods that could be used in the determination of the optimum system can lead to involved discussions of the relative analytical and computational merits of the various methods. In order to avoid these involvements, heuristic arguments are introduced which lead naturally to the explicit determination of both the transient characteristics of the optimum system and the optimum configuration of the system components. In particular, the error criterion appropriate for the design of adaptive feedback controls has been selected by recognizing the role of the conditions existing in the dynamic process (called state of the process) and then assuming that this state could be measured. An effort to preserve a thread of unity dictates that these considerations should be reflected by the mathematical methods used in the explicit optimization. In order to emphasize concepts and not generality, the concepts are identified here by discussing the simplest possible dynamic process which is given by

$$q^{(1)}(\sigma)=b_0(\sigma)m(\sigma) \tag{8}$$

The error e associated with the dynamic process described by equation 8 is given by equation 4. If the assumption is made that the multiplier $\lambda(\sigma)$ and the desired response signal $Q(\sigma)$ are given, this error is a function of the present state of the dynamic process which now is merely $q(t)$, the time t where this process state variable is measured, and whatever control signal $m(\sigma)$ that is inserted into the error criterion. If a dummy variable μ is defined as the time where the state of the dynamic process is measured, the error given in equation 4 can be expressed functionally as

$$e[q(\mu),m(\sigma);\mu]=\int_{\mu}^{t+\tau}\times\{\lambda(\sigma)[Q(\sigma)-q(\sigma)]^2+m^2(\sigma)\}d\sigma \tag{9}$$

However, when equation 9 is minimized, the control signal $m(\sigma)$ is no longer arbitrary but is specified as a function of the remaining independent variable $q(\mu)$. Symbolically, the minimum error E can be written as

$$E=\min_{\substack{m(\sigma)\\ \mu\leq\sigma\leq t+\tau}}\{e[q(\mu),m(\sigma);\mu]\} \tag{10}$$

where the minimization is performed with respect to the control signal $m(\sigma)$ over the interval $\mu\leq\sigma\leq t+\tau$. For minimum error, the control signal is given functionally as

$$m(\sigma)=F_\sigma[q(\mu)] \tag{11}$$

where the dummy variable σ is treated as a parameter in the function F. Therefore equation 10 is rewritten as

$$E=e[q(\mu),F_\sigma[q(\mu)];\mu], \tag{12}$$

and the minimum error is seen to be a function only of the measured state of the dynamic process $q(\mu)$ and the time of measurement μ. Finally then, the minimum error function is defined for the dynamic process described by equation 8 as

$$E[q(\mu);\mu] = \min_{m(\sigma)} \left[\int_{\mu}^{t+\tau} \{\lambda(\sigma)[Q(\sigma) - q(\sigma)]^2 + m^2(\sigma)\} d\sigma \right] \quad (13)$$

$$\mu \leq \sigma \leq t+\tau$$

The reason for defining the minimum error function $E[q(\mu);\mu]$ is that the necessary key step is the use of the functional equation given in equation 11 which gives the optimum control signal. Specifically, equation 11 is a control law for a feedback controller when $\mu = t$ because the optimum control signal $m(t)$ is expressed in terms of the measured response signal $q(t)$. Therefore, the optimum control signal will be found directly when the minimum error function is found without the necessity of converting an optimum transfer function into a feedback system, a procedure associated with the usual calculus of variations. This conversion, of course, is predicted on the assumption of linearity, and hence violates the desire for adaptability. With the use of Bellman's "principle of optimality," the condition for minimum error can be shown to be

$$\min_{m(\mu)} \left\{ \lambda(\mu)[Q(\mu) - q(\mu)]^2 + m^2(\mu) + \frac{\partial E[q(\mu);\mu]}{\partial \mu} + b_0(\mu) m(\mu) \frac{\partial E[q(\mu);\mu]}{\partial q(\mu)} \right\} = 0 \quad (14)$$

where $E[q(t+\tau);\ t+\tau] = 0$ as can be seen from equation 13 with $\mu = t+\tau$.[2,3] The author has omitted the derivation of equation 14 because no additional concepts pertinent to the adaptive control problem per se are required for the derivation. In other words, equation 14 states that the sum of the terms in the braces vanishes when the value of the control signal $m(\mu)$, treated here as a parameter, is adjusted so that this sum is a minimum.

The sum in the braces of equation 14 is minimum when

$$\frac{\partial}{\partial m(\mu)} \left\{ \lambda(\mu)[Q(\mu) - q(\mu)]^2 + m^2(\mu) + \frac{\partial E[q(\mu);\mu]}{\partial \mu} + b_0(\mu) m(\mu) \times \frac{\partial E[q(\mu);\mu]}{\partial q(\mu)} \right\} = 0 \quad (15)$$

which is satisfied when

$$m(\mu) = -\frac{b_0(\mu)}{2} \frac{\partial E[q(\mu);\mu]}{\partial q(\mu)} \quad (16)$$

Substitution of the value of $m(\mu)$ given in equation 16 into equation 14 gives the condition required of the minimum error function for the dynamic process specified by equation 8 as

$$\lambda(\mu)[Q(\mu) - q(\mu)]^2 + \frac{\partial E[q(\mu);\mu]}{\partial \mu} - \left\{ \frac{b_0(\mu)}{2} \frac{\partial E[q(\mu);\mu]}{\partial q(\mu)} \right\}^2 = 0 \quad (17)$$

As noted previously, the feedback control law is given directly in terms of the minimum error function as expressed in equation 16 when $\mu = t$.

Having obtained the condition that specifies the minimum error function in terms of a partial differential equation, this function is sought in explicit form. As would be the case for an unfamiliar ordinary differential equation, a power series solution could be assumed. Then the coefficients in the series could be found by direct substitution.

The solution of equation 17 is assumed to be of the form

$$E[q(\mu);\mu] = K(\mu) + K_0(\mu) q(\mu) + K_{00}(\mu) q^2(\mu) \quad (18)$$

where the K's are the parameters to be determined. The series given in equation 18 is truncated at the squared term because higher order terms can be shown to be zero by the procedure that follows when applied to the more general series. Now remembering that both $q(\mu)$ and μ are treated as independent variables in the definition of the minimum error function, the partial derivatives needed for equation 17 are

$$\frac{\partial E[q(\mu);\mu]}{\partial \mu} = K^{(1)}(\mu) + K_0^{(1)}(\mu) q(\mu) + K_{00}^{(1)}(\mu) q^2(\mu) \quad (19)$$

and

$$\frac{\partial E[q(\mu);\mu]}{\partial q(\mu)} = K_0(\mu) + 2K_{00}(\mu) q(\mu) \quad (20)$$

When equations 19 and 20 are substituted into equation 17, equation 17 can be rewritten by collecting terms according to their power in $q(\mu)$ such that

$$\left[K^{(1)}(\mu) + \lambda(\mu) Q^2(\mu) - \frac{b_0^2(\mu)}{4} K_0^2(\mu) \right] + q(\mu)[K_0^{(1)}(\mu) - 2\lambda(\mu) Q(\mu) - b_0^2(\mu) K_0(\mu) K_{00}(\mu)] + q^2(\mu)[K_{00}^{(1)}(\mu) + \lambda(\mu) - b_0^2(\mu) K_{00}^2(\mu)] = 0 \quad (21)$$

Because three independent parameters K, K_0, and K_{00} appear in the assumed form of the minimum error function, three independent conditions are applied to equation 21. In particular, these parameters are chosen so that the three brackets in equation 21 vanish and, therefore, are given by the three ordinary first-order differential equations

$$K^{(1)}(\mu) = -\lambda(\mu) Q^2(\mu) + \frac{b_0^2(\mu)}{4} K_0^2(\mu) \quad (22)$$

$$K_0^{(1)}(\mu) = 2\lambda(\mu) Q(\mu) + b_0^2(\mu) K_0(\mu) K_{00}(\mu) \quad (23)$$

$$K_{00}^{(1)}(\mu) = -\lambda(\mu) + b_0^2(\mu) K_{00}^2(\mu) \quad (24)$$

These parameters then are specified completely because $K(t+\tau) = K_0(t+\tau) = K_{00}(t+\tau) = 0$, a condition which arises from $E[q(t+\tau);\ t+\tau] = 0$ for all $q(t+\tau)$. Finally, the control law of the optimum controller is given (when $\mu = t$) from equation 16 as

$$m(\mu) = -\frac{b_0(\mu)}{2} K_0(\mu) - b_0(\mu) K_{00}(\mu) q(\mu) \quad (25)$$

However, before a detailed discussion of the properties of the optimum controller is embarked upon, the preceeding development is re-examined for the simplest case of pure-time delay in the dynamic process as given by

$$q^{(1)}(\sigma + T) = b_0(\sigma) m(\sigma) \quad (26)$$

For the error criterion given in equation 7 which applies for pure-time delay, the corresponding minimum error function can be defined as

$$E_T[q(\mu+T);\mu] = \min_{m(\sigma)} \left[\int_{\mu}^{t+\tau'} \{\lambda(\sigma+T) \times [Q(\sigma+T) - q(\sigma+T)]^2 + m^2(\sigma)\} d\sigma \right] \quad (27)$$

$$\mu \leq \sigma \leq t+\tau'$$

such that $E_T[q(t+\tau);\ t+\tau'] = 0$. Here the state of the dynamic process, as measured at time μ in conjunction with an appropriate computation, that specifies the effect of the past on the controllable segment $\mu+T \leq \sigma \leq t+\tau$ of the response signal is given by $q(\mu+T)$.

Corresponding to this minimum error function, the condition similar to equation 14 is

$$\min_{m(\mu)} \left\{ \lambda(\mu+T)[Q(\mu+T) - q(\mu+T)]^2 + m^2(\mu) + \frac{\partial E_T[q(\mu+T);\mu]}{\partial \mu} + b_0(\mu) m(\mu) \frac{\partial E_T[q(\mu+T);\mu]}{\partial q(\mu+T)} \right\} = 0 \quad (28)$$

If the minimum error function is assumed to be of the form

$$E_T[q(\mu+T);\mu] = K(\mu) + K_0(\mu) q(\mu+T) + K_{00}(\mu) q^2(\mu+T) \quad (29)$$

the derivation from equation 15 to equation 24 can be repeated for the pure-time delay case to give the following equations which specify the parameters K, K_0, and K_{00}:

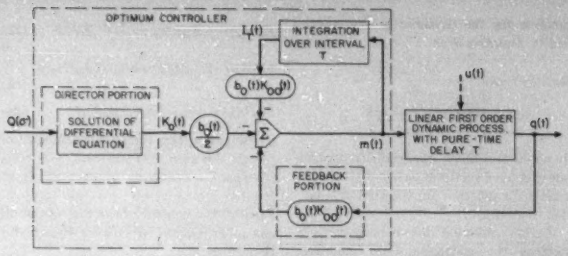


Fig. 3. Block diagram of the optimum controller for a first-order dynamic process with pure-time delay T. Not indicated is that the parameter $K_{\infty}(t)$ also is the solution of a differential equation

$$K^{(1)}(\mu) = -\lambda(\mu+T)Q^2(\mu+T) + \frac{b_0^2(\mu)}{4} K_0{}^2(\mu) \quad (30)$$

$$K_0{}^{(1)}(\mu) = 2\lambda(\mu+T)Q(\mu+T) + b_0{}^2(\mu)K_0(\mu)K_{00}(\mu) \quad (31)$$

$$K_{00}{}^{(1)}(u) = -\lambda(\mu+T) + b_0{}^2(\mu)K_{00}{}^2(\mu) \quad (32)$$

where $K(t+\tau') = K_0(t+\tau') = K_{00}(t+\tau') = 0$. Finally, the control law for the pure-time delay case (when $\mu = t$) which corresponds to equation 25 is

$$m(\mu) = -\frac{b_0(\mu)}{2} K_0(\mu) - b_0(\mu)K_{00}(\mu)q(\mu+T) \quad (33)$$

Properties of the Optimum Adaptive Controller

Having derived the control law for the optimum adaptive controller, a detailed examination of the results indicates some very important properties of the optimum controller from a feedback control standpoint. Because the case with a pure-time delay of duration T degenerates to the case without time delay by allowing $T \to 0$, as can be seen from equation 30 to equation 32 and equation 22 to equation 24, the discussion generally is directed to the case with the pure-time delay. From equation 33 then, the control law for the optimum controller is

$$m(t) = -\frac{b_0(t)}{2} K_0(t) - b_0(t)K_{00}(t)q(t+T) \quad (34)$$

Furthermore, for the dynamic process given in equation 26, the response of the process is

$$q(t) = \int_{-\infty}^{t-T} b_0(\xi)m(\xi)d\xi \quad (35)$$

Therefore, the measured variable $q(t+T)$ that appears in the optimum control law can be written as

$$q(t+T) = q(t) + I_T(t) \quad (36)$$

where

$$I_T(t) = \int_{t-T}^{t} b_0(\xi)m(\xi)d\xi \quad (37)$$

Finally, the optimum control law when written in terms of the measured response signal is

$$m(t) = \left[-\frac{b_0(t)}{2} K_0(t) - b_0(t)K_{00}(t)I_T(t) - b_0(t)K_{00}(t)q(t)\right] \quad (38)$$

which can be depicted as shown in Fig. 3. In general then, the director and feedback portions of the adaptive controller are the result of continually repeated computations of the parameters $K_0(t)$ and $K_{00}(t)$ by the solving of two ordinary differential equations over a time interval τ'. The remaining portion of the controller is the continuous measurement of the integral $I_T(t)$ which is required by the presence of puretime delay.

Now with reference to Fig. 3 and to the steps used in the derivation of the parameters $K_0(t)$ and $K_{00}(t)$ which appear in the block diagram, the important properties of this optimally adaptive controller can be listed as follows.

1. The controller is optimum for all initial conditions $q(t)$ which may occur because the parameters are chosen so that equation 21 is satisfied for all finite $q(t)$ by making each bracket in that equation vanish.

2. The controller is optimum for all future states of the dynamic process, as given by $q(\mu)$ at the interval $t < \mu \leq t+\tau$, from the same reasoning used to establish the previous property. Therefore the controller is optimum for all additive load disturbances (shown dashed in Fig. 3) that cannot be measured and possess the property that their future is independent of their present and past.

3. The feedback portion of the optimum controller, given by $K_{00}(t)$, is independent of the desired response $Q(\sigma)$ as can be seen from equation 32.

4. The output of the director portion of the optimum controller, given by $K_0(t)$, is linear with respect to additive components in the desired response signal $Q(\sigma)$. This property can be seen from equation 31 and results because the parameter $K_{00}(\mu)$ is merely a time varying multiplier, as it is not dependent on $Q(\sigma)$.

The first two properties of the optimum control system give rise to the concept of an optimum configuration. Because the feedback portion of the optimum control system is linear, the feedback loop, for instance, could be replaced by an equivalent linear compensation as indicated in Fig. 4. The system shown in Fig. 4 has the same response signal as the system shown in Fig. 3 when the initial conditions and load disturbances are neglected. However, the configuration shown in Fig. 3 gives minimum error due to both initial conditions and additive load disturbances of the class defined previously, and hence is termed the "optimum configuration." In other words, deviation from the optimum configuration in general causes increased error.

The second two properties of the optimum control system give rise naturally to the suggestion that the controller may have important properties for desired response signals that can be represented by a sum of components. In particular, here the assumption is made that the desired response signal is

$$Q(\sigma) = \sum_{n=0}^{N} f_n(\sigma, t)Q_n(t) \quad (39)$$

where the signal $Q_n(t)$ is a component which can be extracted from $Q(t)$ by a physically realizable device. The necessity for a physically realizable device is stressed here because the segment of the

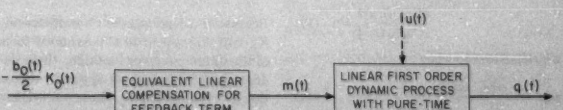


Fig. 4. Block diagram of equivalent linear compensation which is optimum only for zero initial conditions and additive load disturbances

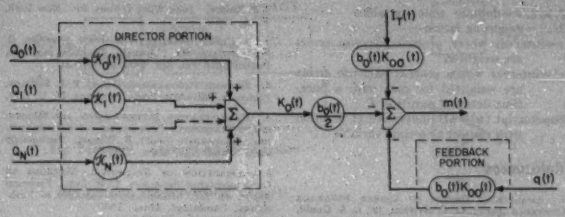


Fig. 5. Block diagram of the optimum linearly-adaptive controller that occurs when the desired response is a sum of separable components

desired response signal that is required for the solution of equation 31 is on an interval of future time $t+T \leq \sigma \leq T+\tau$. As an example in this class of signals, the desired response signal may be approximated with accuracy by a Taylor series such that $Q_n(t) = Q^{(n)}(t)$, and hence the function f_n is $f_n(\sigma,t) = (\sigma-t)^n/n!$ In this example, the component extraction is accomplished by successive differentiation. Now because superposition applies to equation 31, the director output can be written in terms of the integral

$$K_0(t) = \int_t^{t+\tau'} h_T(\tau', t+\tau'-\mu) Q(\mu+T) d\mu \tag{40}$$

The impulse response function $h_T(\tau',\xi)$ is defined as the signal $K_0(\xi)$ at $\zeta=\tau'$ caused by a unit impulse at $\zeta=\xi$ where the dummy variable ζ is defined as $\zeta = t+\tau'-\mu$ for use in equation 31. For the class of desired response signals given by equation 39, the director output signal becomes

$$K_0(t) = \sum_{n=0}^{N} \kappa_n(t) Q_n(t) \tag{41}$$

where the no-energy storage gain $\kappa_n(t)$ is

$$\kappa_n(t) = \int_t^{t+\tau'} h_T(\tau', t+\tau'-\mu) f_n(\mu+T, t) d\mu \tag{42}$$

Therefore these gains are the solution of the differential equation

$$\kappa_n{}^{(1)}{}_{(\mu)} = 2\lambda(\mu+T) f_n(\mu+T, t) + b_0{}^2(\mu) \kappa_n(\mu) K_{00}(\mu) \tag{43}$$

Fig. 5 shows that the optimum controller consists of only time varying gains, for desired response signals of the form of equation 39, and hence the optimum control system is "linearly-adaptive" to the desired response signal.

Conclusions

The goals of this paper have been to demonstate, via the simplest problem, how the selection of an appropriate error criterion and the use of a particular method of optimization apply to the design of a class of adaptive controllers that enjoy the attributes of feedback. In particular, the design problem is restricted to a linear but time-varying first-order dynamic process that may contain a pure-time delay and to an error criterion of quadratic form. For these mathematical restrictions, the optimum controller is shown to be independent of all finite initial conditions and additive load disturbances of the class defined previously. Furthermore, the parameters which appear in the optimum controller are specified at each instant of real time t as the solutions of two ordinary first-order differential equations. Finally, for these mathematical restrictions and a class of desired response signals which are a sum of separable components, the optimum control system is linear but in general time varying by virtue of the presence of time varying gains.

For the purposes of application to engineering design problems, a wide variety of transient characteristics can be obtained by the proper selection of the weighting of response errors, as given by $\lambda(\sigma)$ and τ which are identified with the problem statement. Furthermore, these linearly-adaptive controllers are particularly practical because the differential equations can be solved by either analog or digital computers at a rate high enough for many applications and because the time-varying gains are readily obtained componentry.

Fortunately the properties described here can be extended to more general problems but with varying degrees of complication, depending upon the mathematical restrictions imposed. Some important cases are:

1. For a Uth order linear dynamic process and a quadratic error criterion, the feedback portion of the control system is specified by U parameters which occur in U feedback loops. However, the director is still given by one parameter. The controller parameters are specified from the solution of $2U+U!/2!(U-2)!$ simultaneous first order differential equations. The system is optimum for all initial conditions and all additive load disturbances of the restricted class defined previously. For a desired response signal of the separable class given in equation 39, the optimum system is linearly-adaptive because the director is specified by time varying gains.

2. For the restrictions of case 1 but allowing the desired response signal to have random components, minimum mean-square error in the control and response signals (the errors being caused by errors that occur in the prediction of the desired response signal over the interval of future time where response errors are weighted in the error criterion) gives rise to Wiener prediction of the desired response signal. The optimum mean-square prediction of the response signal has the separable form of equation 39 for linear lumped parameter Wiener predictors, and hence the system is linearly-adaptive as the director is again given by time-varying gains. Here each component signal $Q_n(t)$ is measured with a Wiener filter.[4,5]

3. For certain types of nonlinear dynamic processes of Uth order such as the case where the control signal is limited in range due to saturation but where the process is otherwise linear, a quadratic form of the error criterion gives rise to an optimum control system that has U feedback loops and is specified by $1+U$ parameters which are computed from $2U+U!/2!(U-2)!$ simultaneous ordinary first-order differential equations. However, the system is optimum only for the particular initial condition, additive load disturbance, and member of the ensemble of possible desired response signals for which the parameters are computed. Hence, the optimum controller is nonlinear for all classes of desired response signals. Also because an iterative procedure is required to select the proper solutions for the parameters, application to problems of various time scales generally is more limited.

4. For nonquadratic forms of the error criterion, optimization by continual computations is possible but very lengthy when load disturbances and the like are included. Hence, without the use of "quasi-optimum" techniques now being investigated, this procedure seems to have a very limited application to problems such as may arise in the chemical processing industry.

Nomenclature

$a_n(t)$, $b_n(t)$ = parameters of the linear dynamic process
ϵ = error criterion
$f_n(\sigma, t)$ = function associated with the extrapolation of a component of the desired response signal, $Q_n(t)$
$E[g(\mu); \mu]$ = minimum error function for first-order dynamic process
$h_T(\tau', t+\tau'-\mu)$ = impulse response of the director portion of the adaptive controller
$I_T(t)$ = integral of the control signal over an interval equal to the duration of the time delay

$K(\mu)$, $K_N(\mu)$, $K_M(\mu)$ = parameters of the minimum error function
$\epsilon_n(t)$ = no energy storage gain computed from the director equations
$m(t)$ = control signal of the dynamic process
n = dummy index of summation
$q(t)$ = response signal of the dynamic process
$Q(t)$ = desired process response signal
$Q_n(t)$ = component of the desired response signal
t = present time
T = duration of pure-time delay
u = dummy index of summation
U = order of the dynamic process
ξ, μ, ξ, σ = dummy time variables
$\lambda(\sigma)$ = weighting factor
τ = interval where process response errors are weighted
τ' = interval where process response errors are weighted for case with a pure-time delay of duration T
Superscript (u) = uth derivative of function with respect to a time variable

References

1. Analytical Design of Linear Feedback Controls (book), G. C. Newton, Jr., L. A. Gould, J. F. Kaiser. John Wiley & Sons, Inc., New York, N. Y., 1957.

2. A Class of Optimum Control Systems, C. W. Merriam, III. *Journal*, Franklin Institute, Philadelphia, Pa., Apr. 1959, pp. 267–81.

3. Dynamic Programming (book), R. E. Bellman. Princeton University Press, Princeton, N. J., 1957.

4. Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications (book), N. Wiener. Technology Press, Cambridge, Mass., 1949.

5. Application of Statistical Methods to Communication Problems, Y. W. Lee. *Technical Report no. 181*, Massachusetts Institute of Technology, Cambridge, Mass., 1950.

A Method for the Symbolic Representation and Analysis of Linear Periodic Feedback Systems

EDWARD O. GILBERT
NONMEMBER AIEE

LINEAR FEEDBACK systems with periodic sampling, finite pulsing, finite pulse clamping and similar variations have been a subject of increasing interest in recent years. Applications in such important areas as digital control and time-division multiplexing have motivated continued research and development. This is particularly true for certain types of systems where the theory is extensive and well developed. Unfortunately, there are many types of systems which have received little or no attention.

It is the purpose of this paper to present methods of symbolism and analysis which handle an extensive class of new problems and present an improved treatment of many problems previously considered.

The sampled-data system is the most studied and well-known system type. Here the periodic variation appears as instantaneous periodic sampling of signals. Mathematically, this sampling is represented by multiplying signals by a periodic impulse train. The literature is extensive and includes multirate and cyclic-rate systems.[1–5] The most important tools of analysis are the Z-transform and modified Z-transform.

Farmanfarma has considered finite pulsed systems where the periodic variation is signal multiplication by a train of periodic pulses with finite width and amplitude.[6–8] The P-transform is defined and used in his analysis. Analysis of closed-loop systems is considerably more complicated than that of open-loop systems. Approximate methods, based on sampled-data models have been proposed by Kranc,[9] Murphy and Kennedy,[10] and Tou.[3]

Another type of periodic variation is the operation of finite pulse clamping illustrated in Fig. 1, where Fig. 1(A) shows the signal input $e(t)$ and Fig. 1(B) shows the finite pulse-clamped signal $e_c(t)$. Mathematically,

$$e_c(t) = e(t),\ nT < t \leqslant nT + t_1$$
$$= e(nT + t_1),\ nT + t_1 < t \leqslant (n+1)T \qquad (1)$$

where T is the fundamental period of variation and n ranges over all integers. Analysis techniques for systems with finite pulse-clamped error signals have been developed.[11]

A periodic variation which includes finite pulsing as a special case is the piecewise constant variation of parameters. Such variation can be in gains and/or time constants of control elements. An example of a parameter with a 2-interval variation is the gain:

$$K(t) = K_1,\ nT < t < nT + t_1$$
$$= K_2,\ nT + t_1 < t < (n+1)T \qquad (2)$$

where T is the fundamental period. Finite pulsing obviously occurs when $K_2 = 0$. Unforced systems of this type are conveniently treated by the matrix method of Pipes.[12] The sinusoidal response of electric networks with piecewise constant variation of parameters was treated by Bennett[13] and extended by Desoer.[14] A still more complete theory of such systems has been developed by the author.[11]

All of these variations can be included in a single system. Fig. 2 shows an example of such a system where the error is finite pulse clamped, modified by a digital computer with clamper, and fed into a plant ($p = d/dt$) with periodic gain variation. The finite pulse-clamped error e_c is defined by equation 1 and the variable gain K by equation 2. The digital computer and clamper are described by the equation:

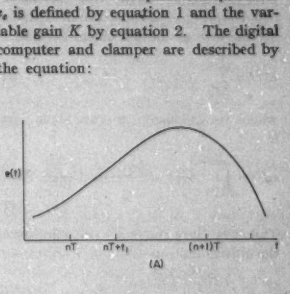


Fig. 1. Operation of finite pulse clamping

A—Continuous signal
B—Finite pulse-clamped signal

Paper 59-1183, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted March 24, 1959; made available for printing September 3, 1959.

Edward O. Gilbert is with the University of Michigan, Ann Arbor, Mich., and on leave of absence with Space Technology Laboratories Inc., Los Angeles, Calif.

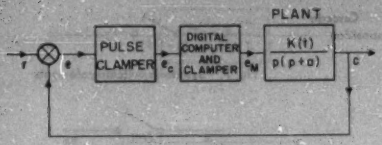


Fig. 2 (above). Example feedback system

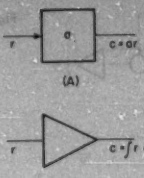


Fig. 3 (right). Basic symbolic elements

A—Multiplication by a constant
B—Integration
C—Summation

$$e_M(t)=\sum_{i=0}^{3} a_i e_c[t_k+(n-i-1)T]-\sum_{j=1}^{3} b_j e_M[t_k+(n-j)T]$$

$$nT<t\leq(n+1)T,\ 0<t_k<T \quad (3)$$

representing instantaneous sampling and clamping.

A typical problem in analysis would be to determine the step response of this system.

The methods of symbolism and analysis presented in this paper treat systems with any combination of the described periodic variations. The formulation is based on a state vector that completely defines the system and input behavior at all times. This state vector is defined by a series of constant coefficient differential equations and transition equations. The equations are readily obtained, from a block diagram consisting of basic symbolic elements. State vectors have been used with success in other applications by Kalman and Bertram,[14,15] Bashkow,[17] and others. The block diagram methods are in some ways similar to those of Bertram.[15]

Solution of the equations is accomplished by matrix methods. The Z-transformation is used to obtain the solution at the fundamental time intervals nT. The continuous solution in any fundamental period $nT<t\leq(n+1)T$ is also obtained. The methods are illustrated with examples.

System Description

All the periodic operations already described are easily represented in block diagram form by five basic operations: multiplication by a constant, integration, summation, switching, and sampling. When every system component is described by symbols representing the basic operations, the mathematical representation in terms of a state vector readily follows.

Invariant Components

Consider first components described by transfer functions with constant coefficients. In this case switching or sampling operations are not necessary and the component can be represented by the block diagram symbols of Fig. 3. The decomposition which results is by no means unique.

To illustrate several decompositions, consider the transfer function:

$$\frac{c}{r}=\frac{p^3+3p+2}{p^3+4.5p^2+2p}=\frac{(p+1)(p+2)}{p(p+0.5)(p+4)} \quad (4)$$

A direct decomposition is shown in Fig. 4. The correctness of this block diagram is made apparent by writing c as:

$$c=(p^{-1}+3p^{-3}+2p^{-3})\bar{c}$$

where

$$\bar{c}=\frac{1}{1+4.5p^{-1}+2p^{-2}}r$$

The state vector is defined by the components y_1, y_2, and y_3, equal in number to the system order. The differential equations for the state variables are readily obtained by inspection. For example, the time rate of change of y_3 equals the sum of inputs to the y_3 integrator, or:

$$\dot{y}_3=-4.5y_3-2y+r$$

Similarly, $\dot{y}_1=y_3$, and $\dot{y}_2=y_3$.

The desired output $c(t)$ is given by a linear combination of the state variables. Thus:

$$c(t)=2y_1+3y_2+y_3$$

By writing the transfer function 4 as the partial fraction expansion

$$\frac{c}{r}=\frac{1}{p}+\frac{-3/7}{p+0.5}+\frac{3/7}{p+4}$$

the block diagram of Fig. 5 results. The state variables are again the integrator outputs defined by the equations:

$$\dot{y}_1=r$$

$$\dot{y}_2=-0.5y+r$$

$$\dot{y}_3=-4y_3+r$$

The output c is given by:

$$c=y_1-\frac{3}{7}y_2+\frac{3}{7}y_3$$

Still another decomposition is obtained by cascading elements of the form $1/(p+a)$ and $(p+b)/(p+c)$. One such form for transfer function 4 is shown in Fig. 6. In this case the state variables are defined by:

$$\dot{y}_1=(1-0.5)y_1+(2-4)y_2+r$$

$$\dot{y}_2=-0.5y_2+(2-4)y_3+r$$

$$\dot{y}_3=-4y_3+r$$

and $c=y_1$.

These and other possible decompositions illustrate the plurality of the symbolic representations. The appropriate decomposition and corresponding state vector are determined by a number of factors. If a transfer function is of high

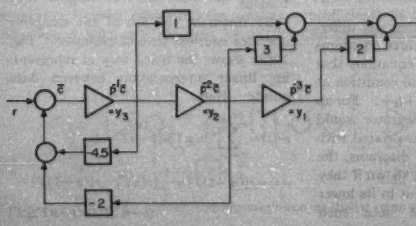


Fig. 4 (left). Direct decomposition

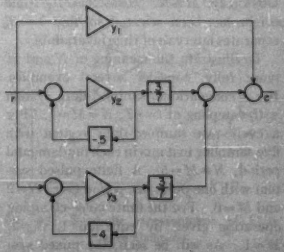


Fig. 5 (right). Partial fraction decomposition

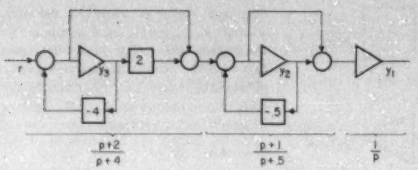


Fig. 6. Cascade decomposition

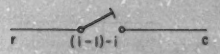
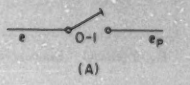


Fig. 7. Symbolic switching elements



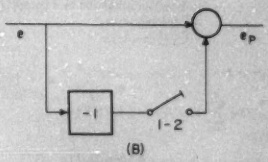


Fig. 8. Two representations of finite pulsing

order and unfactored, the direct decomposition method is most convenient. The partial fraction representation has a computational advantage in that the differential equation for each state variable is independent of the other state variables. Often it is desirable from the standpoint of synthesis and compensation to have the state variables correspond directly to physical quantities. In control systems, this can frequently be done by using the cascade representation.

In many cases the application of physical laws leads directly to a state vector. One example occurs in electrical network analysis when dependent variables are selected as the voltages across capacitors and the currents through inductors.[17] Another example is the linearized equations for an aircraft which are a set of first-order differential equations. In these cases, the state variables represent physical quantities and transfer function derivation is superfluous.

Switching and Sampling

Before describing the basic switching and sampling operations, the interval number and sampling number of a system will be defined. The interval number N is the number of intervals of time invariant behavior in the fundamental period T. These intervals of invariance are specified by $0<t-nT\leq t_1$, $t_1<t-nT\leq t_2$, ..., $t_{n-1}<t-nT\leq t_N=T$. The sampling number M is the number of instantaneous samples in the fundamental period T. Since sampling is a time variability, the sampling instants separate time invariant behavior and must occur at one or more of the values nT, $nT+t_1$, ..., $nT+t_{n-1}$. Obviously, $M\leq N$. Note that the time origin has been selected so that $t=nT$ separates intervals of time invariability.

To illustrate the meaning of N and M more fully, consider several examples. For a conventional sampled-data system with sampling at $t=nT$, $N=M=1$. For a cyclic-rate sampled-data system with five sampling instants in each fundamental period, $N=M=5$. A finite pulsed system with one pulse per period gives $N=2$ and $M=0$. For the finite pulse-clamping operation given by equation 1, $N=2$, $M=1$. As will be seen, the mixed system of Fig. 2 can have several N, M combinations depending on the formulation chosen. It is sometimes useful to think of time invariant systems as a special case where $N=1$, $M=0$.

The $(i-1)-i$ switching operation, shown schematically in Fig. 7 by a bar switch, is defined by:

$$\begin{aligned} c&=0,\ 0<t-nT\leq t_{i-1}\\ &=r,\ t_{i-1}<t-nT\leq t_i\\ &=0,\ t_i<t-nT\leq T \end{aligned} \tag{5}$$

Although N switching operators exist, they may not all be used. An example of this is shown in Figs. 8(A) and (B) where the pulsed error:

$$\begin{aligned} e_p(t)&=e(t),\ 0<t-nT\leq t_1\\ &=0,\ t_1<t-nT\leq T \end{aligned}$$

is obtained in two different ways. Depending on the application, one formulation may be preferable to another.

The switching operations describe any piecewise constant parameter variation. The only other necessary operation is that of sampling. The sampling operator considered here differs from the usual impulse train multiplication. The new method of representation is illustrated in Fig. 9(A) for a 3-sample system where sampling occurs at $t=nT$, $nT+t_1$, and $nT+t_2$. The sampling operation, which always occurs in conjunction with integration is represented by arrow switches, as opposed to the bar switches, feeding the lower side of an integrator. The integrator equations in Fig. 9(A) are:

$$\begin{aligned} y_1&=r_1\\ y_1(nT+)&=r(nT)\\ y_1(nT+t_1+)&=y_1(nT+t_1)\\ y_1(nT+t_2+)&=y_1(nT+t_2) \end{aligned} \tag{6}$$

The first equation is the integrator differential equation; the last three equations are the reset or transition equations that specify the integrator initial condition at $t=nT+$, $nT+t_1+$, and $nT+t_2+$. For an M sample system each integrator would have M sample switches associated with it. To simplify the block diagrams, the sample switches will not be shown if they connect an integrator output to its lower side. This is reasonable since such

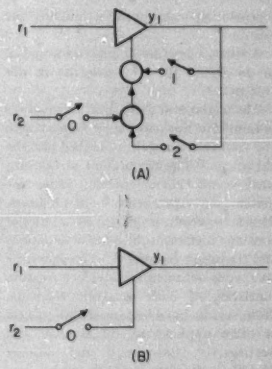


Fig. 9. Symbolic sampling elements

A—Complete form
B—Simplified form

operations give continuity in the integrator variable. Fig. 9(B) shows this simplification for the system of Fig. 9(A). Other examples are illustrated in Figs. 3, 4, 5, and 6 for invariant systems where the integrator outputs are obviously continuous.

To illustrate the use of the sampling operator consider several examples. Fig. 10(A) shows the usual way of representing linear interpolation between data points. Here:

$$r^*(t)=\sum_{n=-\infty}^{\infty} r(nT)\delta(t-nT)$$

$$c(t)=r[(n-1)T]+\frac{1}{T}\{r(nT)-r[(n-1)T]\}$$
$$(t-nT),\ 0<t-nT\leq T$$

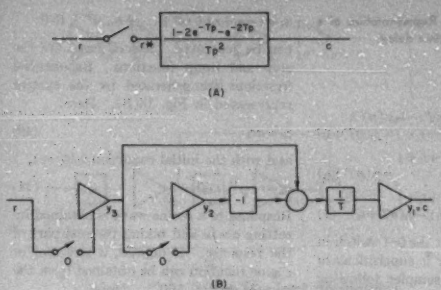


Fig. 10. Linear interpolation device

A—Conventional sampled-data system representation
B—New representation

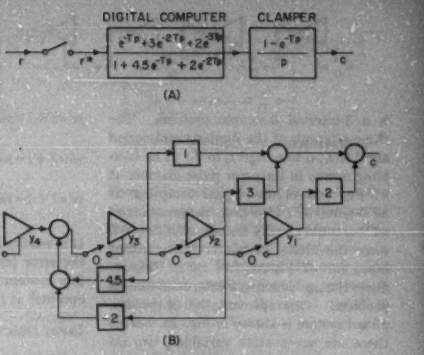


Fig. 11. Digital computer and clamper

A—Conventional sampled-data system representation
B—New representation

The new description, using the basic symbolism, is shown in Fig. 10(B). The variable y_3 is the sampled and clamped $r(t)$; y_2 is the sampled and clamped signal delayed by the period T; y_1 is $1/T$ $\{r(nt)-r[(n-1)T]\}$ for $0<t-nT<T$; and therefore y_1 equals the desired $c(t)$ if $y_1(nT)=r[(n-1)T]$, which is the case if the system is initially at rest (say at $t=0$). The state vector formed by y_1, y_2, and y_3 completely describes the device. The variables are governed by the differential equations:

$$\dot{y}_1=-\frac{1}{T}y_3+\frac{1}{T}y_2$$

$$\dot{y}_2=0$$

$$\dot{y}_3=0$$

and the reset or transition equations:

$$y_1(nT+)=y_1(nT)$$
$$y_2(nT+)=y_2(nT)$$
$$y_3(nT+)=r(nT)$$

A somewhat more general example of a sampled-data system is the digital computer and clamper shown in the usual notation in Fig. 11(A) and in the new notation in Fig. 11(B). The sampled and clamped r is given by the y_4 integrator while the digital computer program is accomplished in the remainder of the block diagram. Note that the digital computer transfer function is the same as equation 4 with p^{-1} replaced by e^{-Tp}.

For this reason, the diagram is analogous to the direct decomposition of Fig. 4. The only difference is the replacement of integrator inputs by reset operations. Other forms of decomposition are of course possible. Fig. 12 illustates the cascade decomposition analogous to Fig. 6. Other decompositions are also possible; for example, the sampling and clamping accomplished in y_4 could take place following the computer representation instead of preceding it. The equations for the state variables of Fig. 11(B) or Fig. 12 are easily written. For the Fig. 12 representation, the differential equations are:

$$\dot{y}_1=0$$

$$\dot{y}_2=0$$

$$\dot{y}_3=0$$

$$\dot{y}_4=0$$

and the reset or transition equations are:

$$y_1(nT+)=(1-0.5)y_3(nT)+(2-4)y_2(nT)+y_4(nT)$$

$$y_2(nT+)=-0.5y_3(nT)+(2-4)y_2(nT)+y_4(nT)$$

$$y_3(nT+)=-4y_2(nT)+y_4(nT)$$

$$y_4(nT+)=r(nT)$$

These transition equations have the same coefficients as the differential equations for the system of Fig. 6.

To represent the finite pulse-clamping operation defined in equation 1, both the switching and sampling operations are necessary. This is obvious from Fig. 13 where the 0–1 and the 1–2 interval equations for e_c are:

$$e_c(t)=e(t),\ 0<t-nT\leq t_1$$
$$=y_1,\ t_1<t-nT\leq T$$

with $\dot{y}_1=0$ and

$$y_1(nT+)=y_1(nT)$$
$$y_1(nT+t_1+)=e(nT+t_1)$$

Time delay is a dynamic characteristic that can be represented when it occurs in a path containing sampling. As an example, suppose that a sampler and a clamper are followed by a time delay of $2T+t_1$, where $0<t_1<T$. The symbolic representation is shown in Fig. 14. Here y_4 is the sampled and clamped r where it is assumed sampling occurs at $t=nT$. The y_2 and y_3 integrators provide a delay of $2T$ while the remaining delay t_1 occurs in the y_1 integrator.

This completes the symbolic description of components necessary to represent more complex systems such as the one in Fig. 2. As described earlier, this system

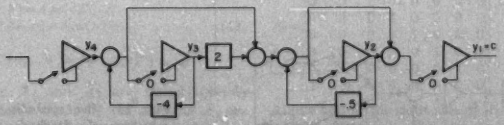


Fig. 12. Alternative representation of digital computer and clamper

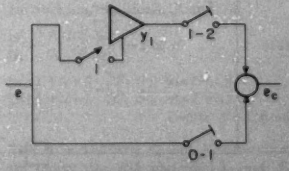


Fig. 13. Symbolic representation of finite pulse clamping

Fig. 14. Representation of a time delay

is a 3-interval 3-sample problem. The three intervals of the fundamental period are 0 to t_1, t_1 to t_2, and t_2 to T. The sampling occurs in the error pulse clamper at $nT+t_1$, and in the digital computer at $nT+t_2$ and nT. However, the sampling at $nT+t_2$ could equally well occur at $(n+1)T$ since the digital computer has constant input for $nT+t_1<t\leq(n+1)T$. This reduces the problem to a 2-interval 2-sample problem. One representation of the simplified system is shown in Fig. 15. Since there are seven state variables, two intervals, and two samples times, fourteen constant coefficient differential equations and fourteen transition equations must be written. The differential equations in interval 0–1 are:

$$\dot{y}_1 = y_2$$

$$\dot{y}_2 = -a_1y_2 + K_1[a_4y_3 + (a_2 - b_2a_0)y_4 + (a_1 - b_1a_0)y_5 + a_0y_6] \quad (7)$$

$$\dot{y}_3 = \dot{y}_4 = \dot{y}_5 = \dot{y}_6 = \dot{y}_7 = 0$$

and in interval 1–2 are:

$$\dot{y}_1 = y_2$$

$$\dot{y}_2 = -a_1y_2 + K_2[a_4y_3 + (a_3 - b_3a_0)y_4 + (a_2 - b_2a_0)y_5 + a_0y_6] \quad (8)$$

$$\dot{y}_3 = \dot{y}_4 = \dot{y}_5 = \dot{y}_6 = \dot{y}_7 = 0$$

The transition equations at $t=nT$ and $t=nT+t_1$ are:

$$y_1(nT+) = y_1(nT),\ y_1(nT+t_1+) = y_1(nT+t_1)$$

$$y_2(nT+) = y_2(nT),\ y_2(nT+t_1+) = y_2(nT+t_1)$$

$$y_3(nT+) = y_4(nT),\quad y_3(nT+t_1+) = y_4(nT+t_1)$$

$$y_4(nT+) = y_3(nT),\ y_4(nT+t_1+) = y_3(nT+t_1)$$

$$y_5(nT+) = y_6(nT) - b_1y_5(nT) - b_2y_4(nT),\quad y_5(nT+t_1+) = y_5(nT+t_1)$$

$$y_6(nT+) = y_7(nT),\ y_6(nT+t_1+) = y_6(nT+t_1)$$

$$y_7(nT+) = y_7(nT),\ y_7(nT+t_1+) = x_1(nT+t_1) - y_1(nT+t_1) \quad (9)$$

It is worthy to note that the 0–1 switch in the error pulse clamper is superfluous in this system since the sampler following operates at $t=nT$ only.

Input Description

Up to now, input variables have been unspecified in functional form. For analysis, it will be useful to think of input quantities as additional state variables described by differential equations and, possibly, transition equations. For a deterministic input such as a step function, ramp function, general polynomial in t, exponential, sinusoid, or certain periodic functions, the formulation of the input state vector offers no particular difficulty. Fortunately, such inputs occur in many important response and synthesis problems. Methods for obtaining the input state variables will now be discussed.

Polynomial inputs are easily generated by the series of integrators shown in Fig. 16(A). Here:

$$\dot{z}_1 = z_2$$
$$\dot{z}_2 = z_3$$
$$\vdots$$
$$\dot{z}_{n-1} = z_n$$
$$\dot{z}_n = 0$$

With the general set of initial conditions $x_1(0)$ $x_2(0)$, ..., $x_{n-1}(0)$, $x_n(0)$ any function of the type

$$x_1 = r = r_0 + r_1t + r_2t^2 + \ldots + r_{n-1}t^{n-1},\ t>0$$

can be generated. Special cases are the step and ramp functions. Exponential functions are generated by the system represented in Fig. 16(B). Here:

$$\dot{x}_1 = ax_1 \quad (10)$$

and with the initial condition $x_1(0)=1$,

$$x_1 = r = e^{at},\ t>0 \quad (11)$$

Response to a cosine wave is obtained by setting $a=j\omega$ and taking the real part of the response. If desired, a sine and/or cosine function can be obtained from the system in Fig. 16(C). Here:

$$\dot{x}_1 = \omega x_2$$
$$\dot{x}_2 = -\omega x_1$$

and with the initial conditions $x_1(0)$ and $x_2(0)$:

$$r = x_1(0)\cos\omega t + x_2(0)\sin\omega t$$

Other functions that are similar or combinations of the above are readily obtained.

More general inputs are obtained using sampling and/or switching. As one example consider the symbolic diagram in Fig. 17 defining a 1-interval 1-sample system. The differential equations are:

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = 0$$
$$\dot{x}_3 = 0$$

the transition equations are:

$$x_1(nT+) = x_1(nT)$$
$$x_2(nT+) = x_2(nT)$$
$$x_3(nT+) = x_3(nT)$$

and the initial conditions at $t=0$ are:

$$x_1(0) = 0$$
$$x_2(0) = \frac{1}{T}$$

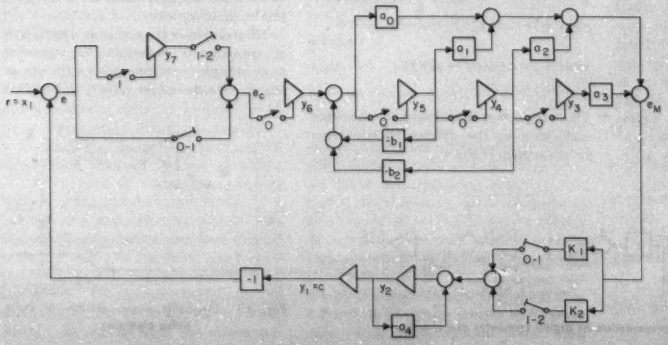


Fig. 15. The symbolic representation of feedback system example

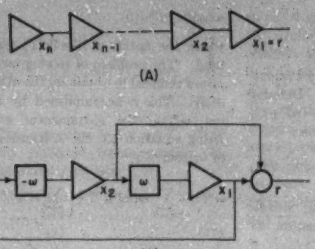


Fig. 16. Input representation

A—Polynomial B—Exponent C—Sinusoid

$x_8(0)=0$

For this system:

$$x_1=r=\frac{t-nT}{T},\ 0<t-nT\leq T,\ t>0$$

This is the equation for a sawtooth wave of unit amplitude and period T. Other periodic inputs, such as a square wave or parabolic wave, are obtainable through similar methods. These more general periodic inputs are limited in that their period or an integral multiple of it must equal the system fundamental period T. It is also possible to generate aperiodic functions of a more general class than defined.

System Equations in Vector Form

By defining the input state vector in the described manner, it is now possible to state concisely equations governing system response. Assume that the system vector $\mathbf{y}$ has k components $y_1, y_2, \ldots, y_k$ and that the input vector $\mathbf{x}$ has m components $x_1, x_2, \ldots, x_m$. Furthermore, assume that the input and system has N invariant intervals and M sampling instants in the fundamental period T. If $\mathbf{w}$ is defined as the vector with the $m+k$ components $x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_k$, the system is described by the N vector differential equations:

$$\begin{aligned}&\dot{\mathbf{w}}=A_1\mathbf{w},\ 0<t-nT\leq t_1\\&\dot{\mathbf{w}}=A_2\mathbf{w},\ t_1<t-nT\leq t_2\\&\quad\vdots\\&\dot{\mathbf{w}}=A_N\mathbf{w},\ t_{N-1}<t-nT\leq T\end{aligned}\qquad(12)$$

and the N vector transition equations:

$$\begin{aligned}&\mathbf{w}(nT+)=\bar{A}_0\mathbf{w}(nT)\\&\mathbf{w}(nT+t_1+)=\bar{A}_1\mathbf{w}(nT+t_1)\\&\quad\vdots\\&\mathbf{w}(nT+t_{N-1}+)=\bar{A}_{N-1}\mathbf{w}(nT+t_{N-1})\end{aligned}\qquad(13)$$

where A_i and $\bar{A}_i$ are $(m+k)\times(m+k)$ matrices. $N-M$ of the matrices $\bar{A}_i$ are equal to the identity matrix I while the remaining M matrices $\bar{A}_i$ are not. Equations 12 and 13 and an initial condition on $\mathbf{w}$ at some time completely specifies the system response.

The matrices A_i and $\bar{A}_i$ are easily obtained from the input and system state variable differential and transition equations. Consider the system in Fig. 15 forced by the exponential of Fig. 16(B). The differential and transition equations are given by 7, 8, 9, 10, and 11. From these the four necessary 8×8 matrices are:

$$A_1=\begin{array}{c}\begin{array}{cccccccc}x_1&y_1&y_2&y_3&y_4&y_5&y_6&y_7\end{array}\\\left\|\begin{array}{cccccccc}a&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&-a_4&K_3a_3&K_3(a_2-b_3a_3)&K_3(a_1-b_3a_2)&K_1a_4&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\end{array}\right\|\end{array}\begin{array}{c}\dot{x}_1\\\dot{y}_1\\\dot{y}_2\\\dot{y}_3\\\dot{y}_4\\\dot{y}_5\\\dot{y}_6\\\dot{y}_7\end{array}$$

$$A_2=\left\|\begin{array}{cccccccc}a&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&-a_4&K_3a_3&K_3(a_2-b_3a_3)&K_3(a_1-b_3a_2)&K_1a_4&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\end{array}\right\|\qquad(14)$$

$$\bar{A}_0=\left\|\begin{array}{cccccccc}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&1\\0&0&0&0&0&0&0&1\end{array}\right\|$$

$$\bar{A}_1=\left\|\begin{array}{cccccccc}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&1&0&0&0&0\\0&0&0&0&1&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\1&-1&0&0&0&0&0&0\end{array}\right\|$$

Above and to the right side of A_1 are the variables corresponding to the respective columns and rows of the matrix.

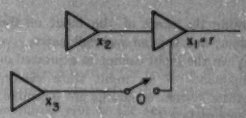


Fig. 17. Sawtooth wave representation

From this example, it is apparent that the matrices can be written by inspection of the system and input symbolic diagrams. A particular response problem is specified by an initial condition on $\mathbf{w}$. As an example, suppose that the system is initially at rest and the input $x_1=e^{at}$ is applied at $t=0$. Then

$$\mathbf{w}(0)=\left\|\begin{array}{c}x_1(0)\\y_1(0)\\y_2(0)\\y_3(0)\\y_4(0)\\y_5(0)\\y_6(0)\\y_7(0)\end{array}\right\|=\left\|\begin{array}{c}1\\0\\0\\0\\0\\0\\0\\0\end{array}\right\|$$

In this example many of the matrix components are zero. Frequently, but not always, this is the case. However, the i, j components of the A and $\bar{A}$ matrices for $i=1$ to m and $j=m+1$ to $m+k$ are always zero, since the output components y_i cannot influence the input components x_i. This is seen by partitioning equations 12 and 13:

$$\left\|\begin{array}{c}\dot{\mathbf{x}}\\\dot{\mathbf{y}}\end{array}\right\|=\left\|\begin{array}{cc}F_i&0\\G_i&S_i\end{array}\right\|\left\|\begin{array}{c}\mathbf{x}\\\mathbf{y}\end{array}\right\|,\ i=1\text{ to }N\qquad(15)$$

$$\left\|\begin{array}{c}\mathbf{x}(nT+t_i+)\\\mathbf{y}(nT+t_i+)\end{array}\right\|=\left\|\begin{array}{cc}\bar{F}_i&0\\\bar{G}_i&\bar{S}_i\end{array}\right\|\left\|\begin{array}{c}\mathbf{x}(nT+t_i)\\\mathbf{y}(nT+t_i)\end{array}\right\|$$

$i=0$ to $N-1$, $t_0=0$

The $m\times m$ matrices F_i and $\bar{F}_i$ determine the input; the $k\times k$ matrices S_i and $\bar{S}_i$ define the system dynamics; the $k\times m$ matrices G_i and $\bar{G}_i$ are gain factors in-

dicating the effect of the input on the system; and 0 is an $m \times k$ null matrix.

When the input cannot be expressed in components of an input state vector, separate system equations are desirable. From equation 15.

$$\dot{\mathbf{y}} = S_i \mathbf{y} + G_i \mathbf{x}, \; i = 1 \text{ to } N$$

$$\mathbf{y}(nT + t_i +) = \bar{S}_i \mathbf{y}(nT + t_i) + \bar{G}_i \mathbf{x}(nT + t_i), \; i = 0 \text{ to } N-1 \quad (16)$$

where $\mathbf{x}$ is now a scalar input.

System Response

Basic Techniques

In the previous section an extensive class of periodic system response problems have been concisely formulated in the $2N$ differential and transition equations 12 and 13. The solution of these equations will now be obtained.

The basic procedure is to work from interval to interval piecing the solution together. Start at $t = nT$ where the solution, yet unknown, is $\mathbf{w}(nT)$. Application of the first transition equation gives:

$$\mathbf{w}(nT+) = \bar{A}_0 \mathbf{w}(nT) \quad (17)$$

This serves as the initial condition on the first differential equation of 12 that defines the solution for $0 < t - nT \leq t_1$. Thus the vector equation $d\mathbf{w}/dt = A_1 \mathbf{w}$ must be solved with the initial condition $\mathbf{w}(nT+)$. This is conveniently accomplished by the Laplace transform. First define the new time scale $\tau = t - nT$ translating $t = nT$ into $\tau = 0$. The problem is now stated:

$$\frac{d\mathbf{w}}{d\tau} = A_1 \mathbf{w}, \; \mathbf{w}(\tau)\big|_{\tau=0} = \mathbf{w}(nT+) \quad (18)$$

Using the notion $\bar{\mathbf{w}}(s) = L[\mathbf{w}(\tau)]$ to indicate the Laplace transform of $\mathbf{w}(\tau)$ with respect to τ, the transform of equation 18 is:

$$s\bar{\mathbf{w}}(s) - \mathbf{w}(nT+) = A_1 \bar{\mathbf{w}}(s) \quad (19)$$

Solving for $\bar{\mathbf{w}}$ and using the inverse Laplace transform gives:

$$\mathbf{w}(\tau) = L^{-1}[(sI - A_1)^{-1}]\mathbf{w}(nT+) \quad (20)$$

The matrix function $L^{-1}[(sI - A_1)^{-1}]$ deserves special notation: define the matrix time function:

$$\epsilon^{A_1 \tau} = L^{-1}[(sI - A_1)^{-1}] \quad (21)$$

This function and methods for simplifying its computation are discussed in the Appendix. The solution for $0 < t - nT \leq t_1$, is now completed by substituting for τ in equation 21, using equations 20 and 17. The result is:

$$\mathbf{w}(t) = \epsilon^{A_1(t-nT)} \bar{A}_0 \mathbf{w}(nT), \; 0 < t - nT \leq t_1 \quad (22)$$

A similar procedure is used in the second interval.

First:

$$\mathbf{w}(nT + t_1 +) = \bar{A}_1 \mathbf{w}(nT + t_1) \quad (23)$$

where $\mathbf{w}(nT + t_1)$ is obtained by evaluating equation 22 at $t = nT + t_1$. Defining another new time scale $\tau = t - (nT + t_1)$, the second equation of 12 becomes:

$$\frac{d\mathbf{w}}{d\tau} = A_2 \mathbf{w}, \; \mathbf{w}(\tau)\big|_{\tau=0} = \mathbf{w}(nT + t_1 +) \quad (24)$$

Use of the Laplace transform, substitution for τ, and use of equation 23 yields:

$$\mathbf{w}(t) = \epsilon^{A_2(t - nT - t_1)} \bar{A}_1 \mathbf{w}(nT + t_1), \quad t_1 < t - nT \leq t_2 \quad (25)$$

where

$$\epsilon^{A_2 \tau} = L^{-1}[(sI - A_2)^{-1}] \quad (26)$$

Substituting in equation 25 for $w(nT + t_1)$ from equation 22 gives:

$$\mathbf{w}(t) = \epsilon^{A_2(t - nT - t_1)} \bar{A}_1 \epsilon^{A_1 t_1} \bar{A}_0 \mathbf{w}(nT), \quad t_1 < t - nT \leq t_2 \quad (27)$$

The solution for the remaining N-2 intervals of $0 < t - nT \leq T$ is obtained in a similar manner.

The solution derived is expressed more concisely by defining the matrix time function:

$$W(\tau) = \epsilon^{A_1 \tau} \bar{A}_0, \; 0 < \tau \leq t_1$$
$$\vdots$$
$$= \epsilon^{A_i(\tau - t_{i-1})} \bar{A}_{i-1} \epsilon^{A_{i-1}(t_{i-1} - t_{i-2})} \cdots \bar{A}_0, \; t_{i-1} < \tau \leq t_i \quad (28)$$
$$\vdots$$
$$= \epsilon^{A_N(\tau - t_{N-1})} \bar{A}_{N-1} \cdots \bar{A}_0, \quad t_{N-1} < \tau \leq t_N = T$$

where of course

$$\epsilon^{A_i \tau} = L^{-1}[(sI - A_i)^{-1}] \quad (29)$$

If $M < N$, $W(\tau)$ is simpler in form since $N - M$ of the $\bar{A}_i$ equal the identity matrix. The solution $\mathbf{w}(t)$ is now easily written as:

$$\mathbf{w}(t) = W(t - nT)\mathbf{w}(nT), \; 0 < t - nT \leq T \quad (30)$$

This is the equation for the solution interior to any fundamental interval.

The first step in obtaining the solution at multiples of T is to evaluate equation 30 for $t = (n+1)T$ giving the vector difference equation:

$$\mathbf{w}[(n+1)T] = W(T)\mathbf{w}(nT) \quad (31)$$

Since this equation is valid for all n, start with $n = 0$ and the known initial condition $\mathbf{w}(0)$. After n applications of equation 31, it is found that:

$$\mathbf{w}(nT) = W^n \mathbf{w}(0) \quad (32)$$

where the notation $W = W(T)$ has been used. The problem of solving for $\mathbf{w}(nT)$ is now reduced to obtaining the nth power of W. This is accomplished by obtaining $\mathbf{w}^*(z)$, the Z-transform of $\mathbf{w}(t)$. Using equation 32, the Z-transform can of written:

$$\mathbf{w}^*(z) = \sum_{n=0}^{\infty} \mathbf{w}(nT) z^{-n} = \left[\sum_{n=0}^{\infty} W^n z^{-n} \right] \mathbf{w}(0) \quad (33)$$

where it is understood that $W^0 = I$. But the bracketed infinite matrix series is recognized as the matrix function $[I - z^{-1}W]^{-1}$.[18] Thus:

$$\mathbf{w}^*(z) = [I - z^{-1}W]^{-1} \mathbf{w}(0) \quad (34)$$

is the closed form of the Z-transform. Application of the inverse Z-transform to equation 34 gives the desired values $\mathbf{w}(nT)$. Alternatively, the components of equation 34, ratios of polynomials in z^{-1}, can be divided out and the coefficient of z^{-n} equated to the components of $\mathbf{w}(nT)$.

As an illustration of the method, consider the response of a first-order system to a suddenly applied sawtooth wave. The transfer function and input are defined by

$$\frac{c}{r} = \frac{1}{\frac{1}{a}p + 1}$$

and

$$r = 0, \; t < 0$$
$$= \frac{1}{T}(t - nT), \; 0 < t - nT \leq T$$

Initially the system is at rest so $c(0) = 0$. Fig. 18 shows the symbolic representation of the input, identical to that of Fig. 17, and the system. By inspection of the diagram, the two required matrices:

$$A_1 = \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a & 0 & 0 & -a \end{Vmatrix} \quad (35)$$

$$\bar{A}_0 = \begin{Vmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

are obtained. Since $N = M = 1$, post-multiplying $\epsilon^{A_1 t}$, which is evaluated in the Appendix, by $\bar{A}_0$ gives the desired:

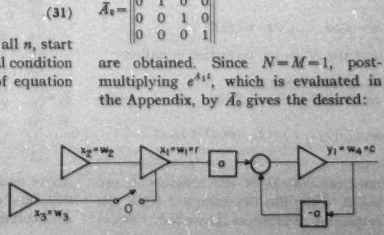


Fig. 18. Example of input and the system

$$W(\tau)=\begin{Vmatrix} 0 & \tau & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \tau-\frac{1}{a}(1-e^{-a\tau}) & (1-e^{-a\tau}) & e^{-a\tau} \end{Vmatrix} \tag{36}$$

Evaluating equation 36 for $\tau=T$, $\mathbf{w}^*(z)$ is readily obtained from equation 34 using the initial condition:

$$\mathbf{w}(0)=\begin{vmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \\ y_1(0) \end{vmatrix}=\begin{vmatrix} 0 \\ \frac{1}{T} \\ 0 \\ 0 \end{vmatrix} \tag{37}$$

Because of the zero components in $\mathbf{w}(0)$, only the second columns of $(I-z^{-1}W)^{-1}$ need be computed. Multiplication of this second column by $1/T$ then gives

$$\mathbf{w}^*(z)=\begin{vmatrix} \frac{z^{-1}}{1-z^{-1}} \\ \frac{1}{T} \\ \frac{1}{1-z^{-1}} \\ 0 \\ \left[1-\frac{1}{Ta}(1-e^{-aT})\right]z^{-1} \\ \hline (1-z^{-1})(1-e^{-aT}z^{-1}) \end{vmatrix} \tag{38}$$

Evaluating the inverse Z-transform of each component at $t=nT$ yields the desired solution:

$$\mathbf{w}(nT)=\begin{vmatrix} u(nT-T) \\ \frac{1}{T}u(nT) \\ 0 \\ u(nT)\left(\frac{1}{1-e^{-aT}}-\frac{1}{aT}\right)(1-e^{-anT}) \end{vmatrix} \tag{39}$$

where the step function $u(t)$ is zero for $t<0$ and unity for $t\geq 0$. The first three components, $x_1(nT)$, $x_2(nT)$, and $x_3(nT)$, obviously agree with specified input state vector. System response is given by the fourth component, $w_4(nT)=y_1(nT)=c(nT)$. From equations 30 and 39 the response $w_4(t)=c(t)$ for $t>0$ is:

$$c(t)=\frac{t-nT}{T}-\frac{1}{aT}+\left[\frac{1}{aT}+\left(\frac{1}{1-e^{-aT}}-\frac{1}{aT}\right)(1-e^{-anT})\right]e^{-a(t-nT)},\quad 0<t-nT\leq T \tag{40}$$

Note that the input and output values $x_2(nT)$ and $y_1(nT)$ were needed to obtain equation 40.

Detailed Solution for N = M = 2

The general method just presented has the advantage of notational simplicity. A more detailed form of the solution with some computational advantage can be obtained using the partitioned equations 15. Unfortunately, equations are excessively long and burdensome to derive if N is large. For this reason only the $N=M=2$ case will be considered here. The solution for higher N and M would proceed in a similar manner. Actually many problems fall into $N=M=2$ class.

To obtain the desired solution, it is necessary to express $e^{At\tau}$ and $W(\tau)$ in partitioned form. From equations 29 and 15:

$$e^{At\tau}=L^{-1}\left[\begin{Vmatrix} (sI-F_i) & 0 \\ -G_i & (sI-S_i) \end{Vmatrix}^{-1}\right]$$
$$=L^{-1}\left[\begin{Vmatrix} (sI-F_i)^{-1} & 0 \\ (sI-S_i)^{-1}G_i(sI-F_i)^{-1} & (sI-S_i)^{-1} \end{Vmatrix}\right] \tag{41}$$

This is expressed more simply as:

$$e^{At\tau}=\begin{Vmatrix} e^{F_i\tau} & 0 \\ G_i(\tau) & e^{S_i\tau} \end{Vmatrix} \tag{42}$$

by defining

$$e^{F_i\tau}=L^{-1}[(sI-F_i)^{-1}]$$
$$e^{S_i\tau}=L^{-1}[(sI-S_i)^{-1}]$$
$$G_i(\tau)=L^{-1}[(sI-S_i)^{-1}G_i(sI-F_i)^{-1}] \tag{43}$$

where $i=1$ and 2. Substituting equation 42 and the partitioned transition matrices of 15, equation 28 then gives:

$$W(\tau)=\begin{Vmatrix} X(\tau) & 0 \\ G(\tau) & Y(\tau) \end{Vmatrix} \tag{44}$$

where

$$X(\tau)=e^{F_1\tau}F_0,\ 0<\tau\leq t_1$$
$$=e^{F_2(\tau-t_1)}F_1e^{F_1t_1}F_0,\ t_1<\tau\leq T$$
$$Y(\tau)=e^{S_1\tau}S_0,\ 0<\tau\leq t_1$$
$$=e^{S_2(\tau-t_1)}S_1e^{S_1t_1}S_0,\ t_1<\tau\leq T$$
$$G(\tau)=G_1(\tau)F_0+e^{S_1\tau}G_0,\ 0<\tau\leq t_1 \tag{45}$$
$$=[G_2(\tau-t_1)F_1+e^{S_2(\tau-t_1)}G_1]e^{F_1t_1}F_0+$$
$$e^{S_2(\tau-t_1)}S_1[G_1(t_1)F_0+e^{S_1t_1}G_0],\ t_1<\tau\leq T$$

The solution for $t=nT$ is obtained by the Z-transform as before. By defining $X=X(T)$, $Y=Y(T)$, and $G=G(T)$, equation 34 becomes

$$\mathbf{w}^*(z)=\begin{vmatrix} \mathbf{x}^*(z) \\ \mathbf{y}^*(z) \end{vmatrix}=\begin{Vmatrix} (I-z^{-1}X) & 0 \\ -z^{-1}G & (I-z^{-1}Y) \end{Vmatrix}^{-1}\begin{vmatrix} \mathbf{x}(0) \\ \mathbf{y}(0) \end{vmatrix}$$
$$=\begin{Vmatrix} (I-z^{-1}X)^{-1} & 0 \\ z^{-1}(I-z^{-1}Y)^{-1}G(I-z^{-1}X)^{-1} & (I-z^{-1}Y)^{-1} \end{Vmatrix}\begin{vmatrix} \mathbf{x}(0) \\ \mathbf{y}(0) \end{vmatrix} \tag{46}$$

From this

$$\mathbf{y}^*(z)=z^{-1}(I-z^{-1}Y)^{-1}G\mathbf{x}^*(z)+(I-z^{-1}Y)^{-1}\mathbf{y}(0) \tag{47}$$

where

$$\mathbf{x}^*(z)=(I-z^{-1}X)^{-1}\mathbf{x}(0) \tag{48}$$

The forced part of the solution 47 is the first term; the part of the solution due to an initial $\mathbf{y}(0)$ is the second term, obviously zero for a system initially at rest. This equation for $\mathbf{y}^*(z)$ is correct in form for any N and M. Higher N and M only result in increased complexity of the equations analogous to equation 45.

After $\mathbf{y}(nT)$ is obtained from equation 47 by the inverse Z-transform or component division, equations 30 and 44 can be used to write:

$$\mathbf{y}(t)=Y(t-nT)\mathbf{y}(nT)+G(t-nT)\mathbf{x}(nT),\quad 0<t-nT\leq T \tag{49}$$

Note that the input vector at $t=nT$ is required.

The complete solution described by equations 47, 48, and 49 is simple in form. For complicated systems, determination of $X(\tau)$, $Y(\tau)$, $G(\tau)$, and the Z-transforms may be lengthy, but the work is straightforward and systematic. To illustrate the method and allow comparison with other methods, when other methods apply, several examples will now be discussed.

Examples

Example I

Fig. 19 shows the symbolic representation for a sampled-data feedback system where the sampled error is clamped and fed into the transfer function $G(p)=K/p(p+a)$. This is an $N=M=1$ system and the results for the $N=M=2$ system apply by taking $t_1=T$ and $S_1=I$. The system matrices are:

$$S_1=\begin{Vmatrix} 0 & 1 & 0 \\ 0 & -a & K \\ 0 & 0 & 0 \end{Vmatrix} \tag{50}$$

$$S_0=\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{Vmatrix}$$

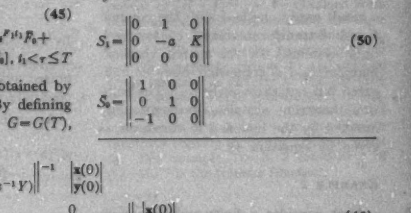


Fig. 19. Sampled-data system example

Now

$$e^{\mathcal{S}_1 T} = \begin{Vmatrix} 1 & \frac{1}{a}(1-\epsilon^{-aT}) & \frac{K}{a}\left[T-\frac{1}{a}(1-e^{-aT})\right] \\ 0 & \epsilon^{-aT} & \frac{K}{a}(1-\epsilon^{-aT}) \\ 0 & 0 & 1 \end{Vmatrix} \tag{51}$$

and thus

$$Y = Y(T) = \epsilon^{\mathcal{S}_1 T}\mathcal{S}_0 = \begin{Vmatrix} \left\{1-\frac{K}{a}\left[T-\frac{1}{a}(1-\epsilon^{-aT})\right]\right\} & \left\{\frac{1}{a}(1-\epsilon^{-aT})0\right\} \\ -\frac{K}{a}(1-\epsilon^{-aT}) & \epsilon^{-aT} \quad 0 \\ -1 & 0 \quad 0 \end{Vmatrix} \tag{52}$$

Since $t_1 = T$, equation 45 shows that $G = G(T) = G_1(T) + \epsilon^{\mathcal{S}_1 T}\mathcal{G}_0$ for $\mathcal{F}_0 = I$. But $G_1 = 0$ and

$$\mathcal{G}_0 = \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix} \tag{53}$$

Thus:

$$G = \begin{vmatrix} \frac{K}{a}\left[T-\frac{1}{a}(1-\epsilon^{-aT})\right] \\ \frac{K}{a}(1-\epsilon^{-aT}) \\ 1 \end{vmatrix} \tag{54}$$

In this case G is the same for any input, so for all $r^*(z) = x_1^*(z)$, $y^*(z) = z^{-1}(I - z^{-1}Y)^{-1}Gx_1^*(z)$ where it is assumed that system is initially at rest. Computation of the component $y_1^* = c^*(z)$ gives:

$$c^*(z) = \frac{\frac{K}{a^2}\{(aT-1+\epsilon^{-aT})z^{-1} + [1-(aT+1)\epsilon^{-aT}]z^{-2}\}r^*(z)}{1-\left[1-\frac{K}{a^2}(aT-1)+\left(1-\frac{K}{a^2}\right)\epsilon^{-aT}\right]z^{-1}+\left[\frac{K}{a^2}+\left(1-\frac{K}{a^2}-\frac{KT}{a}\right)\epsilon^{-aT}\right]z^{-2}} \tag{55}$$

a result easily checked with the conventional Z-transform theory of sampled-data systems. In the matrix method, manipulation is somewhat more complicated but not excessively so. The solution between samples, conventionally handled by the modified Z-transform, is given by equation 49.

Example 2

Next consider a finite pulsed system analyzed by Farmanfarma.[8] In this system the plant input is $e = r - c$ for $0 < t - n < 0.6$ and zero for $0.6 < t - n < 1$. The plant transfer function is $G(p) = 6/p(p+5)$. Fig. 20 shows the symbolic diagram for the system and input when $r = x_1 = \epsilon^{at}$. Since in this case $N = 2$ and $M = 0$, $\mathcal{S}_0 = \mathcal{S}_1 = \mathcal{F}_0 = \mathcal{F}_1 = I$ and $\mathcal{G}_0 = \mathcal{G}_1 = 0$. By inspection:

$$\mathcal{S}_1 = \begin{Vmatrix} 0 & 1 \\ -6 & -5 \end{Vmatrix}, \quad G_1 = \begin{vmatrix} 0 \\ 6 \end{vmatrix}$$

$$\mathcal{S}_0 = \begin{Vmatrix} 0 & 1 \\ 0 & -5 \end{Vmatrix}, \quad G_0 = \begin{vmatrix} 0 \\ 0 \end{vmatrix} \tag{56}$$

From equations 10 and 11, desired input is obtained when $F_1 = F_2 = a$. After obtaining $\epsilon^{\mathcal{A}_1 T}$ and $\epsilon^{\mathcal{A}_2 T}$ and computing $Y = \epsilon^{\mathcal{A}_1 0.4}\epsilon^{\mathcal{A}_2 0.6}$, it is possible to find.

$$(I - z^{-1}Y)^{-1} = \begin{Vmatrix} \frac{1+0.0144z^{-1}}{1-0.417z^{-1}+0.0067z^{-2}} & \frac{0.1175z^{-1}}{1-0.417z^{-1}+0.0067z^{-2}} \\ (\quad) & (\quad) \end{Vmatrix} \tag{57}$$

Only $y_1^*(z) = c^*(z)$ is of interest so the second row is not computed.

If stability is the only question, only the denominator of each component, equal to the $\det(I - z^{-1}Y)$, need be computed. In this case the system is stable since both the z roots are within the unit circle. Bertram and Kalman in a discussion of Farmanfarma's paper[8] discussed this method of stability determination.

To continue with the response problem, G is obtained from equation 45, which in this case is the simpler form:

$$G(\tau) = \epsilon^{\mathcal{S}_0(\tau - t_1)}G_1(t_1)$$

where

$$G_1(\tau) = L^{-1}\left[(sI - \mathcal{S}_1)^{-1}G_1\frac{1}{s-a}\right]$$

Carrying out the computation for $G = G(1)$,

$$G = \frac{1}{\left(\frac{a}{2}+1\right)\left(\frac{a}{3}+1\right)}\begin{vmatrix} -0.1175a - 0.432 + (1+0.173a)\epsilon^{0.6a} \\ 0.0144a - 0.1104 + 0.1353a\epsilon^{0.6a} \end{vmatrix} \tag{58}$$

In this case, G depends on the input indicating the greater complexity of finite pulsed systems as compared with sampled-data systems.

Assuming the system is initially at rest,

$$y_1^*(z) = c^*(z) = \frac{[-0.1175a - 0.432 + (1+0.173a)\epsilon^{0.6a}]z^{-1} + [0.0067 + (0.0144 + 0.0184a)\epsilon^{0.6a}]z^{-2}}{\left(\frac{a}{2}+1\right)\left(\frac{a}{3}+1\right)(1 - 0.417z^{-1} + 0.0067z^{-2})(1 - \epsilon^a z^{-1})} \tag{59}$$

is obtained from equation 47. Setting $a = 0$ and dividing out the polynomial gives $c(n)$ values that check with Farmanfarma's Fig. 10.[8]

Example 3

As a final example consider the finite pulse-clamped system shown in Fig. 21. Take $T = 1$ so that the plant receives the actual error for $0 < t - n \leq t_1$, and the t_1 clamped error for $t_1 < t - n < 1$, where the plant has the transfer function $G(p) = 2/p$. The initial conditions $x_1(0) = 1$, and $y_1(0) = y_2(0) = 0$ specify a step-response problem.

By inspection:

$$\mathcal{S}_1 = \begin{Vmatrix} -2 & 0 \\ 0 & 0 \end{Vmatrix}, \quad \mathcal{S}_0 = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}$$

$$\mathcal{S}_2 = \begin{Vmatrix} 0 & 2 \\ 0 & 0 \end{Vmatrix}, \quad \mathcal{S}_1 = \begin{Vmatrix} 1 & 0 \\ -1 & 0 \end{Vmatrix} \tag{60}$$

$$G_1 = \begin{vmatrix} 2 \\ 0 \end{vmatrix}, \quad \mathcal{G}_0 = \begin{vmatrix} 0 \\ 0 \end{vmatrix}$$

$$G_2 = \begin{vmatrix} 0 \\ 0 \end{vmatrix}, \quad \mathcal{G}_1 = \begin{vmatrix} 0 \\ 1 \end{vmatrix}$$

$$\mathcal{F}_1 = \mathcal{F}_2 = 0, \quad \mathcal{F}_0 = \mathcal{F}_1 = 1$$

After $\epsilon^{\mathcal{A}_1 T}$ and $\epsilon^{\mathcal{A}_2 T}$ are obtained,

$$Y = \epsilon^{\mathcal{A}_2(1-t_1)}\mathcal{A}_1\epsilon^{\mathcal{A}_1 t_1}\mathcal{A}_2 = \begin{Vmatrix} \epsilon^{-2t_1}(2t_1 - 1) & 0 \\ -\epsilon^{-2t_1} & 0 \end{Vmatrix} \tag{61}$$

is readily found. Then:

$$(I - z^{-1}Y)^{-1} = \begin{Vmatrix} \frac{1}{1-\epsilon^{-2t_1}(2t_1-1)z^{-1}} & 0 \\ \frac{-\epsilon^{-2t_1}z^{-1}}{1-\epsilon^{-2t_1}(2t_1-1)z^{-1}} & 1 \end{Vmatrix} \tag{62}$$

Stability is determined by the component denominators. In this case the system is stable for all t_1 except $t_1 = 0$, where neutral stability exists.

By equations 45 and 60,

$$G = G(1) = \epsilon^{\mathcal{S}_0(1-t_1)}\mathcal{G}_1 + \epsilon^{\mathcal{S}_0(1-t_1)}\mathcal{S}_1 G_1(t_1) \tag{63}$$

where by equation 43

$$G_1(\tau) = L^{-1}\left[\begin{Vmatrix} \frac{1}{s+2} & 0 \\ 0 & \frac{1}{s} \end{Vmatrix}\begin{Vmatrix} 2 \\ 0 \end{Vmatrix}\begin{vmatrix} \frac{1}{s} \\ 0 \end{vmatrix}\right] \tag{64}$$

Carrying out the indicated operations:

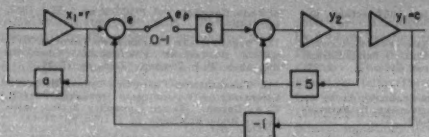


Fig. 20 (above). Finite pulsed system example

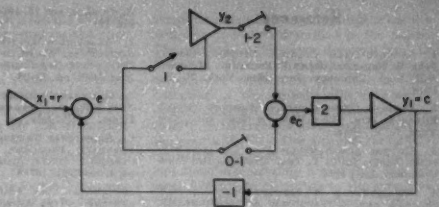


Fig. 21 (right). Finite-pulse clamped-system example

$$G=\left|\frac{1-(2t_1-1)\epsilon^{-2t_1}}{\epsilon^{-2t_1}}\right| \tag{65}$$

Substituting in equation 47 from 62 and 65, the component $y_1^*(s)=c^*(s)$ is given by:

$$y_1^*(s)=c^*(s)$$
$$=\frac{[1-(2t_1-1)\epsilon^{-2t_1}]z^{-1}}{[1-(2t_1-1)\epsilon^{-2t_1}z^{-1}](1-z^{-1})} \tag{66}$$

This result reduces to the sampled-data case when $t_1=0$ and the continuous case when $t_1=1$. Applications of the final value theorem shows that $c(\infty)=1$ as expected. The system has dead-beat response for $t_1=1/2$. More complicated examples of finite pulse-clamped systems are analyzed in a similar way.

Conclusions

This paper presents methods of symbolic representation and analysis applicable to an extensive class of periodic linear systems. The symbolic representations are easily obtained and lead directly to a concise mathematical formulation in terms of a state vector representing both the input and system variables. System response at multiples of the fundamental period, obtained by Z-transform methods, is presented in a concise vector matrix notation. The usual Z-transform methods for stability prediction and final value determination apply. Response interior to the fundamental periods is readily obtained using time matrices and the solution at multiples of the fundamental period.

Appendix

The matrix time function

$$\epsilon^{A\tau}=L^{-1}[(sI-A)^{-1}] \tag{67}$$

results naturally from the vector differential equation and initial condition

$$\frac{dw}{d\tau}=Aw,\ w(0) \tag{68}$$

This is obvious from the Laplace transform of equation 68 which gives:

$$\bar{w}(s)=(sI-A)^{-1}w(0) \tag{69}$$

The computation of $\epsilon^{A\tau}$ by equation 67 is perfectly straightforward but may become lengthy if the matrix order Q is high. One difficulty is the rapidly increasing complexity of matrix inversion with increasing Q; another problem is that the $(sI-A)^{-1}$ components, in general $Q-1$ order polynomials in s divided by Q order polynomials in s, rapidly increase in number and complexity with increasing Q. However, there are frequently many zero entries in the A matrix. As a result, matrix inversion is less complex than expected. Also, $(sI-A)^{-1}$ typically has many components that are zero or simple in form.

In most cases the computation of $\epsilon^{A\tau}$ is greatly simplified by relating the components of $\epsilon^{A\tau}$ to the symbolic diagram. To see this, write:

$$w(\tau)=\epsilon^{A\tau}w(0) \tag{70}$$

obtained by applying the inverse Laplace transform to equation 69 and using equation 67. Assume that the jth component of $w(0)$ is one and that all other components of $w(0)$ are zero. Then by equation 70, the ith component of $w(\tau)$ is the i, j component of $\epsilon^{A\tau}$. In terms of the symbolic diagram, the i, j component of $\epsilon^{A\tau}$ is the ith integrator output for unit initial condition on the jth integrator and zero initial condition on all other integrators. This interpretation of $\epsilon^{A\tau}$ immediately gives all the components of $\epsilon^{A\tau}$ which are zero and frequently allows simple evaluation of the remaining nonzero components.

To illustrate, consider the system in Fig. 18. Since $\epsilon^{A\tau}$ describes the system between sample intervals, neglect the sample switch. Consider first-unit initial value on w_1. By inspection of Fig. 18 i, 1 components, forming the first column of $\epsilon^{A\tau}$, are 1, 0, 0, and $1-\epsilon^{A\tau}$. For unit initial value on w_2, the i, 2 components forming the second column of $\epsilon^{A\tau}$ are τ, 1, 0, and $\tau-1/a\ (1-\epsilon^{-a\tau})$. The other columns are obtained similarly. Thus:

$$\epsilon^{A\tau}=\begin{Vmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1-\epsilon^{-a\tau} & \tau-\frac{1}{a}(1-\epsilon^{-a\tau}) & 0 & \epsilon^{-a\tau} \end{Vmatrix} \tag{71}$$

In more complicated systems, any component of $(sI-A)^{-1}$ can be obtained from the symbolic block diagram using the same basic idea. Thus the i, j component to $(sI-A)^{-1}$ is the transfer function relating the ith integrator output to an artificial input summed in directly at the jth integrator output. This is made clear in Fig. 22 for the system of Fig. 4. Here the $i, 2$ components of $(sI-A)^{-1}$ are the transfer functions relating y_i to the artificial input u_2. For example, the 3, 2 component of $(sI-A)^{-1}$ is the transfer function:

$$\frac{-2s}{s^3+4.5s+2}$$

relating y_3 to u_2.

Another method, particularly useful in approximation, is to use the infinite series representation of $\epsilon^{A\tau}$. By substitution into equations 70 and 68, it is easily shown that the infinite matrix series:

$$\epsilon^{A\tau}=I+A\tau+A^2\frac{\tau^2}{2!}+A^3\frac{\tau^3}{3!}+\dots \tag{72}$$

is valid. When τ is small, two or three terms of this series yield a good approximation to the desired function.

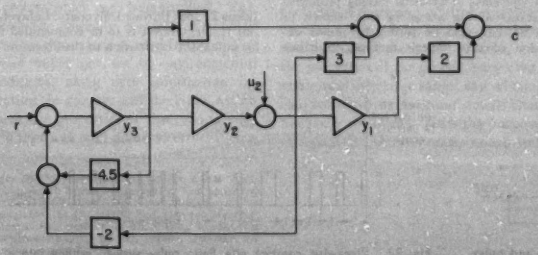


Fig. 22. Symbolic diagram method for determining $(sI-A)^{-1}$ components

References

1. Sampled-Data Control Systems (book). **John R. Ragazzini, Gene F. Franklin.** McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

2. Sampled-Data Control Systems (book). **Eliahu I. Jury.** John Wiley & Sons, Inc., New York, N. Y., 1958.

3. Digital and Sampled-Data Control Systems (book). **Julius T. Tou.** McGraw-Hill Book Company, Inc., 1959.

4. Bibliography of Sampled-Data Control Systems and Z-Transform Applications, **H. Freeman, O. Lowenchuss.** *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-4, Mar. 1958.

5. Analysis of Cyclic-Rate Sampled-Data Feedback-Control Systems, **R. E. Hufnagel.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, Nov. 1958, pp. 421–25.

6. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, **G. Farmanfarma.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, 1956 (Jan. 1957 section), pp. 808–19.

7. Analysis of Multiple Sampler Systems with Finite Pulse Width, Open Loop, **G. Farmanfarma.** *Ibid.*, pt. II (*Applications and Industry*), vol. 77, Mar. 1958, pp. 20–28.

8. General Analysis and Stability Study of Finite Pulsed Feedback Systems, **G. Farmanfarma.** *Ibid.*, July, pp. 148–62.

9. Additional Techniques for Sampled-Data Feedback Problems, **G. M. Kranc.** *Wescon Convention Record*, Institute of Radio Engineers, pt. 4, 1957, pp. 157–65.

10. Closed-Loop Analysis of Sampled-Data Systems with Appreciable Pulse Width, **Gordon J. Murphy, Harold B. Kennedy.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, 1958 (Jan. 1959 section), pp. 659–65.

11. A Theoretical Study of Linear Dynamic Systems with Periodic, Piecewise Constant Parameters, **Edward O. Gilbert.** *Ph.D. Dissertation*, University of Michigan, Ann Arbor, Mich., Jan. 1957.

12. Four Methods for the Analysis of Time-Variable Circuits, **L. A. Pipes.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-2, no. 1, Mar. 1955, pp. 4–12.

13. Steady-State Transmission Through Networks Containing Periodically Operated Switches, **W. R. Bennett.** *Ibid.*, pp. 17–21.

14. Transmission Through a Linear Network Containing a Periodically Operated Switch, **C. A. Desoer.** *Wescon Convention Record*, Institute of Radio Engineers, pt. 2, 1958, pp. 34–41.

15. General Synthesis for Computer Control of Single-Loop and Multiloop Linear Systems (An Optimal Sampling System), **R. E. Kalman, J. E. Bertram.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, 1958 (Jan. 1959 section), pp. 602–09.

16. The Effect of Quantization in Sampled-Feedback Systems, **J. E. Bertram.** *Ibid.*, Sept. 1958, pp. 177–82.

17. The A-Matrix, New Network Description, **T. R. Bashkow.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-4, no. 3, Sept. 1957, pp. 117–19.

18. The Theory of Matrices (book), **C. C. MacDuffee.** Chelsea Publishing Company, New York, N. Y., 1946.

Discussion

E. I. Jury (University of California, Berkeley, Calif.): The author is to be commended for the general analysis of a large class of periodic feedback systems in a unified symbolic representation. This form of representation is very useful in interpreting the steps of the mathematical formulation and is, in some cases, useful in the simulation studies of these systems. Among the important classes this method describes is the case of periodically varying parameters. This includes the category of gain, time constants, pulse duration or the sampling period, and periodic variations. This category is among the topics of recent investigation in this field and this paper appears timely and illuminating.

The analytical solution of the problem is generally based on two steps: 1. the formulation of the difference equations which describe the systems, and 2. the stability study and solutions of these equations for certain periodic or aperiodic inputs or disturbances. For the class of systems mentioned, the difference equations are generally linear with periodic coefficients and, for simpler cases, with constant coefficients.

To obtain the difference equations for these systems, the author has made use of the state vector[1] concept; however, at this point it might be well to mention that one can also arrive at this step by simple manipulation of the ρ-transform and Z-transform methods or by writing the difference equations in time domain if the transfer function is given (or can be easily obtained) in terms of differential equations.[2] For the solution of these difference equations, there also exist several methods, among them the Z-transform[3] method and the matrix form,[4] or a combination of both, as used by the author. The choice among the several methods of formulation and solution depends on the problem and on ease and convenience of arriving at the required results. However, the labor involved is more or less equal among the several methods.

I have recently proposed in a paper written with T. Nishimura the analysis of finite pulsed feedback systems with periodically varying sampling rate and pulse width. Fig. 23 illustrates the basic concepts involved in this problem. Although this case was not treated by the author, the method of the paper is applicable, for it falls under the general category (N, M) with $M=0$ for this case. Our method of approach, based on the ρ-transform and Z-transform methods known for the conventional cases, is quite different from the method of the paper, indicating an alternate approach to analysis of the general class of (N, M) systems. It might be indicated that for the simpler cases of linear sampled-data systems, the basic Z-transform or ρ-transform methods are easily applicable without having to use the general method of the paper. However, it is for larger values of (N, M) that this method becomes more effective and useful.

Furthermore, it is worthwhile to mention that the class of systems having piecewise constant-varying sampling rate[5] shown in Fig. 24, can also be solved by minor extension of this paper's method. Further work on investigating the various sampling schemes as well as the synthesis procedure that could be tackled with this method is indeed warranted.

In conclusion, this paper represents a useful contribution to the growing area of the field and the symbolic method introduced will undoubtedly enhance the methods and techniques available for the formulation and analysis of the general class of linear periodic feedback systems.

References

1. A Unified Approach to the Theory of Sampling Systems, **R. E. Kalman, J. E. Bertram.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 267, no. 5, May 1959, pp. 405–36.

2. The Analysis of Sampled-Data Control Systems with a Periodically Time Varying Sampling Rate, **E. I Jury, F. J. Mullin.** *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-4, no. 1, May 1959.

3. A Note on the Operational Solution of Linear Difference Equations, **E. I. Jury, F. J. Mullin.** *Journal*, Franklin Institute, vol. 266, no. 3, Sept. 1958.

4. Theory of Time Varying Sampled-Data Systems, **B. Friedland.** *Technical Report T-19/B*, Electronics Research Laboratory, Columbia University, New York, N. Y., Apr. 1957.

5. Analysis of Aperiodically-Sampled-Data Feedback Control Systems, **R. E. Hufnagel.** *Ph.D. Dissertation*, Cornell University, Ithaca, N. Y., June 1959.

Julius T. Tou (Purdue University, Lafayette, Ind.): Dr. Gilbert is to be commended for his valuable contribution to the literature of

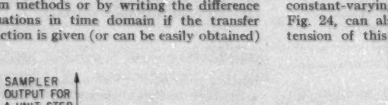


Fig. 23. Periodically time-varying sampling rate and pulse width

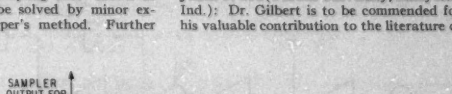


Fig. 24. Piecewise constant rate finite pulse system, settling into a fixed period

sampled-data control systems. The method presented by the author is systematic and straightforward and appears to be a better technique for analyzing finite pulsed feedback control systems by use of Z-transforms. However, it appears that the new symbolic representation by the author does not simplify the problem. On the other hand, the conventional symbolic representation is probably easier and simpler to use. The continuous-data part of the system and the continuous input signals can be represented by the conventional analog-computer simulation diagrams. The pulsed-data part of the system can be described by the conventional symbolic representation given in reference 3 of the paper, with each ideal sampler followed by a zero-order hold. When a sampled-data system is so represented, the state and the transition equations, like equations 8 and 9 of the paper, can readily be obtained.

It would be clearer if $k_0=0$ is added to equation 5 which defines the switching operation analytically. According to the definition of equation 5, Fig. 8(B) can represent a number of switching operations depending upon the value of k_1, and it is equivalent to Fig. 8(A) only if $k_1=T$.

In general, system analysis and synthesis may be carried out by two major approaches. One approach involves the determination of the transfer characteristics of the system components and the over-all transfer characteristic. This approach is usually effected by a block diagram representation and may be referred to as the block diagram approach. The other approach is based upon the characterization of a system by a number of simple first-order differential equations describing the state variables, with the initial conditions given by the transition equations. Each component of a system is decomposed into the basic mathematical elements describing it. This approach is usually effected by a state variable diagram and may be referred to as the state diagram approach.

The author has written a concise and lucid paper on the state diagram approach of the analysis of linear periodic feedback systems.

Edward O. Gilbert: The discussions by Professors Jury and Tou are indeed appreciated. Their discussions and recent papers,[1,2] published or submitted after submission of the author's paper, indicate high interest in more general time-variant operations.

Professor Jury is correct in pointing out the usefulness of the representation in system simulation. There is a 1-to-1 correspondence between the symbolic diagram and differential analyzer setup, including discrete components. Using this analogy, the author has successfully simulated digital computer components by approximating the sampling operation described in the paper. The method of the paper does not lead to difference equations with periodic coefficients as does the method of reference 2. The reduction of the periodic difference equations to constant coefficient difference equations[2,3] is inherent in the computation of the matrix $W(T)$. The ρ-transform and Z-transform methods for writing difference equations are certainly acceptable, but are more difficult to generalize to the extensive class considered in this paper. This, of course, does not invalidate the usefulness of such approaches, which for certain systems, primarily low (N, M) systems, may be more desirable. The same may be said for the various techniques of solving the resulting difference equations.

Professor Tou's statement that the symbolic representation does not simplify the problem and is not necessary to readily derive the state and transition equations is only true for the simplest of systems. For example, the finite pulse-clamping operation is not easily described by conventional representations. The correction in regard to Fig. 8(B) is appreciated.

REFERENCES

1. See reference 1 of the Jury discussion.
2. See reference 2 of the Jury discussion.
3. See reference 3 of the Jury discussion.

Executive-Controlled Adaptive Systems

ROBERT STAFFIN
NONMEMBER AIEE

THERE IS an immense body of literature on the general subject of feedback control theory. The greatest portion of this literature is devoted to the analysis and synthesis of single-loop or simple multiloop linear or piecewise-linear feedback systems. In most of this work, there is a common fundamental assumption, and that is that the designer has, or can obtain, reasonably complete knowledge of the properties of the device or process which is to be controlled. This knowledge generally consists of a fairly accurate characterization of the process or device under one set of environmental conditions, along with information indicating the manner and the limits within which the properties of the process vary with time and with different external conditions. It is the ability of closed-loop control to render the complete system relatively insensitive to inaccuracies in the characterization and to variations in the properties of the process to be controlled that has resulted in the phenonomenal growth of the field.

In this introductory era of supersonic flight, guided missiles, space travel, and complex industrial processes, the control engineer is frequently confronted with the problem of designing control systems for processes where little significant information is known about the process, where the properties of the process vary over an extraordinarily large range, and where the characteristics of the input signals change markedly with time. It is the purpose of this paper to present an approach to the design of feedback control systems which will enable the designer to cope with such situations when they arise.

Paper 59-1181, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted August 3, 1959; made available for printing August 13, 1959.

ROBERT STAFFIN is with Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

Adaptive Systems

A block diagram of a single-loop feedback system is shown in Fig. 1. The compensator can be a passive or an active network, linear or nonlinear, a digital controller, or an analog computer. If the characteristics of the process and of the input signal are known, there are numerous procedures for designing the compensator so that some desired performance is obtained. Deviations from this desired performance due to changes in the process and/or the input signal often can be made acceptably small by increasing the loop gain. If, however, the parameter variations are extremely large, the gain required to achieve the specified system performance may become so high as to be unobtainable because of noise or saturation limitations. As an alternative to increasing the gain, the transfer char-

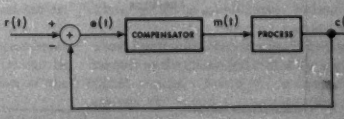


Fig. 1. General single-loop feedback system

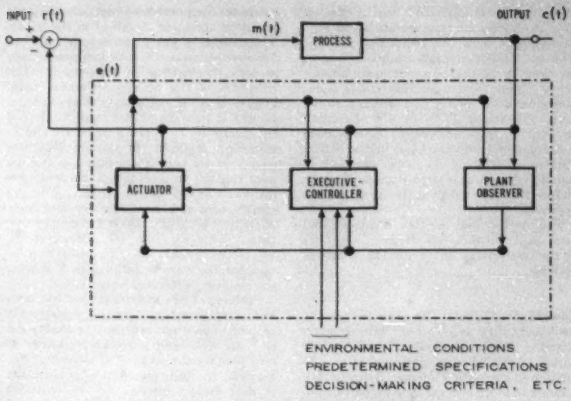


Fig. 2. Executive-controlled adaptive system

acteristic of the compensator can be altered. (The term "transfer characteristic" is used instead of transfer function so that linearity is not implied. If the compensator is linear, the terms are synonymous.) This type of system in which the transfer characteristic of the controller is adjusted to compensate for variations in the characteristics of the signal and/or the process is called an "adaptive system."

In the process-adaptive system[1,2] in which the compensator is designed to adjust for variations in the process, the object is essentially to maintain a relatively constant loop transmittance. In the signal-adaptive system,[2] however, the object is to maintain some specified overall performance in spite of large variations in input-signal characteristics. This latter type of adaptability therefore requires that the compensator vary the loop transmittance as the properties of the input signal change. This fundamental difference in the function of the compensator for process-adaptive and signal-adaptive systems is primarily responsible for the fact that in the past, process adaptability and signal adaptability have been treated as separate design problems.

There is, however, an approach to the design of adaptive systems which enables the designer to develop a system which can be made simultaneously signal-adaptive and process-adaptive. The resulting systems which also can be constructed as either process-adaptive or signal-adaptive are called "executive-controlled adaptive systems."[4,5]

Executive-Controlled Adaptive Systems

In an executive-controlled system, the compensator is a computer which is divided into three basic sections, each with a sharply defined responsibility. An operations diagram for such a system, which may include either an analog, a digital, or a hybrid analog-digital computer, is shown in Fig. 2. The plant observer is responsible for the measurement of the process characteristics and for the detection of changes in significant parameters; the executive-controller is responsible for supplying the required performance information: and the actuator is charged with the computation and generation of the control signal $m(t)$. Although the actual details of each computer section are dependent upon the complexity of the control problem, the following is a general description of the operation of an executive-controlled system.

The actuator, which may range in complexity from an elaborate digital installation to a simple passive network, is designed so that its transfer characteristic is a function both of the significant process characteristics as determined by the plant observer and of the performance requirements as contained in the executive controller. For processes which vary markedly in an extreme and unpredictable manner, the required plant observer may involve a reasonably large computational facility. However, as in the examples which are presented later, a simple computer is often sufficient for handling a number of practical control problems which are commonly encountered. For processes wherein the variations are related to environmental conditions, the plant observer reduces to a set of simple environmental measuring instruments, and, in the limit, for a system which is to be solely signal adaptive, the plant-observer stage is omitted completely.

The executive-controller is responsible for supplying the actuator with performance information. In one of the examples presented later, the executive-controller appears in the form of a set of potentiometer settings. In another, it is a passive network. The actual form of the controller is strongly dependent upon the actuator which is used. As an example, the actuator of Fig. 3 is adjusted to approximately cancel the poles and zeros of the process. For this case, the executive-controller is a model of the desired forward-path transmittance. Although in general, the executive-controller can be called upon to operate on almost any performance criteria, for the purpose of this paper, the executive-controller is used exclusively to achieve a desired forward-path transmittance.

Assuming that the actuator is capable of forcing the forward path to behave as required by the executive-controller, it is

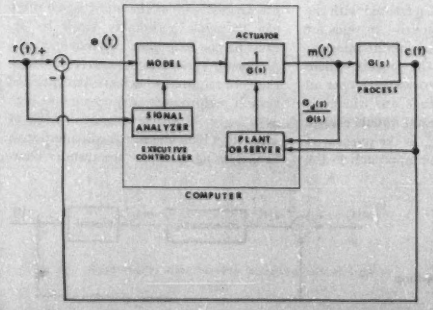


Fig. 3. Function diagram of executive-controller and actuator

apparent that in order to design a signal-adaptive system, it is necessary to vary the parameters of the executive-controller in accordance with the input-signal characteristics. In the system of Fig. 3, this implies that the characteristics of the model are changed as the properties of the input signal vary. For signal-adaptive systems, all signal measuring equipment is included in the executive-controller. Since the problems associated with the measurement of process characteristics are extremely challenging, and because at present, systems which adjust to extreme process variations are of primary interest, the emphasis in the illustrative examples is on process-adaptive executive-controlled systems.

Process Measurement

If the complete executive-controlled system is to adjust automatically to extreme and unpredictable variations in the process characteristics, one section of the computer must be devoted to the continuous measurement and computation of the significant process parameters. The problems associated with this so-called plant-observer stage are enormous. Fundamentally, they fall into two categories: difficulties due to stored energy in the system, and erratic results caused by measurement noise.

Measurements are made by the plant observer over a finite interval of time. During this interval, the output of the process to be measured is not only a function of the input during the interval of measurement, but is also a function of the input for all time prior to the measurement interval. The measurement technique must be capable of taking into account this stored energy which is due to inputs prior to the measurement interval.

Measurement noise is the term applied to errors which arise either as a result of noise within the process itself or from inaccuracies in the measuring equipment. It is apparent that a measured output due to noise in the process with no corresponding input signal would yield nonsensical information about the process. Effects due to measurement noise can be minimized by artifically maintaining high signal levels or by making measurements only when the signals rise above a predetermined threshold value.

The two basic approaches to the solution of these difficulties involve systems which employ test signals as opposed to systems which compute the significant process characteristics from measurements made on the signals present during ordinary operating conditions. Systems which utilize test signals offer two advantages. First, the effects of stored energy due to prior inputs are filtered out. Second, the magnitude of the test signal can be maintained at a level high enough to eliminate most of the errors due to measurement noise. The major disadvantage associated with test-signal systems is, of course, the continual disturbance of the system by the test-signal input.

Systems that compute on the basis of signals which exist during ordinary operation are free from this disturbing feature. However, in order to insure the reliability of the measured data, measurements must be made only when the signals are above some threshold level. An unfortunate consequence of this threshold arrangement is that after periods of low signal levels, the executive-controlled system operates with stale process information. If the occasional lack of immediate information proves more disturbing than test-signal inputs, then the systems which employ test signals are preferable.

Regardless of which approach is chosen, the complexity of the plant-observer stage is dependent upon the number of process parameters which are required in order to appropriately adjust the actuator. For this reason, attempts to solve the process measurement problem in general usually result in enormous computer requirements. Fortunately, for many given practical control situations, the measurement problem can be greatly simplified. The following examples illustrate this point as well as a number of the general chraacteristics of executive-controlled adaptive systems.

Particular Solution to the General Problem

The design of the computer for an executive-controlled system depends quite markedly on the choice of characterization for the process and for the performance specification. This characterization may be in terms of a set of difference equations, by pole-zero location, in terms of an impulse response, or by a frequency spectrum. A number of computer designs based upon these characterizations are discussed in references 1, 2, 4, and 5.

One interesting approach results from choosing a set of linear differential equations to describe the process and the required performance specification. Assuming a linear process, the process input $m(t)$ and the process output $c(t)$ can be related by a differential equation of the form:

$$\sum_{k=0}^{n} a_k p^k c(t) = \sum_{j=0}^{n} b_j p^j m(t) \qquad (1)$$

where $p = d/dt$ and the a_k and b_j are constants. The performance specification which, in this case, is the desired forward-path transmittance, can also be expressed in terms of a differential equation:

$$\sum_{k=0}^{n} \alpha_k p^k c(t) = \sum_{j=0}^{n} \beta_j p^j e(t) \qquad (2)$$

where $c(t)$ is now the desired process output for a given input $r(t)$, and $e(t)$ is the closed-loop error signal $r(t) - c(t)$.

If equation 2 is subtracted from equation 1 and the resultant equation solved for the actuator output $m(t)$, the transfer equation for the actuator is obtained as:

$$m(t) = \frac{1}{b_0} \sum_{k=0}^{n} \left[(a_k - \alpha_k) \frac{d^k}{dt^k} c(t) + \beta_k \frac{d^k}{dt^k} e(t) \right] - \sum_{j=1}^{n} b_j \frac{d^j}{dt^j} m(t) \qquad (3)$$

For the case when $b = 0$, the equation can be solved for $dm(t)/dt$ where $m(t)$ is given to within an additive constant by:

$$m(t) = \frac{1}{b_1} \int \left\{ \sum_{k=0}^{n} \left[(a_k - \alpha_k) \frac{d^k}{dt^k} c(t) + \beta_k \frac{d^k}{dt^k} e(t) \right] - \sum_{j=1}^{n} b_j \frac{d^j}{dt^j} m(t) \right\} dt \qquad (4)$$

For $b_0 = b_1 = 0$, two integrations are required, and so on.

If equations 3 and 4 are normalized for convenience ($-1 < a,\ b,\ \alpha,\ \beta < 1$), one possible computer circuit for generating $m(t)$ utilizing d-c operational amplifiers and single-pole–double-throw switches or relays is shown in Fig. 4(A). The magnitudes of the a_k, β_k, α_k and b_k are set on the potentiometers and the required algebraic signs are obtained by selecting the proper switch position. It is interesting to note that in this method of realization, the executive-controller and the actuator are essentially fused together, the executive-controller taking form as a set of potentiometer settings.

Of extreme importance, however, is the fact that as long as the a_k and b_k correctly characterize the process the forward loop is constrained to perform in the manner described by specification equation 2. Thus, if a signal-adaptive system is desired, it is only necessary to vary the α and β potentiometers in a predetermined manner as the characteristics of the input-signal change. For a process-adaptive system, the a and b potentiometers are

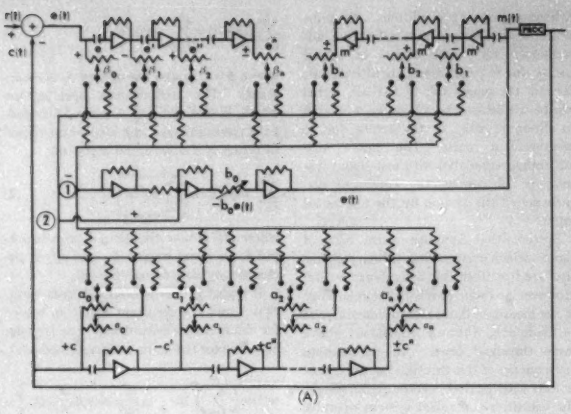


Fig. 4. Differential equation approach

A—Circuit diagram
B—Plant observer

adjusted in accordance with information obtained from a plant observer.

The plant observer for the actuator of Fig. 4(A) is required to supply the coefficients of the process differential equation. If the process input $m(t)$ and output $c(t)$ and all n derivatives of $m(t)$ and $c(t)$ are sampled at $2(n+1)$ instants of time, then process equation 1 can be used to generate a set of linear simultaneous equations. The a_k and b_k are obtained from the solution of this set of equations. A block diagram of the plant observer which operates in this manner is shown in Fig. 4(B). The differentiators can be constructed from analog components, utilizing either operational amplifiers or delay-line differentiators. For the simultaneous equation solver, a digital computer would be more appropriate. The outputs from the equation solver are transformed into potentiometer settings through a positioning servomechanism.

As indicated earlier, the total installation represented by Figs. 4(A) and (B) is quite large, especially for a process which must be represented by a high-order differential equation. In addition, the utilization of a large number of differentiators in an analog computer is usually to be avoided. Fortunately, when the problem is restricted, the instrumentation can be greatly simplified. It is apparent, for example, that for linear systems, the transfer characteristic of the actuator can be writen as $G_d(s)/G(s)$ where $G(s)$ is the process transfer function and $G_d(s)$ is the desired forward-path transmittance. For simple special cases, the computer can therefore be replaced by a network realization of $G_d(s)/G(s)$, even when $G_d(s)$ or $G(s)$ are to be varied during operation.

For systems, however, in which the parameter variations are extreme, the actuator must be realized, either partially or totally as shown in Figs. 4(A) and (B). (For a more detailed analysis of this and other approaches, see reference 5.)

Fortunately, the solution for a particular problem is usually much simpler than that for the general problem just discussed. This simplification is demonstrated in the next section.

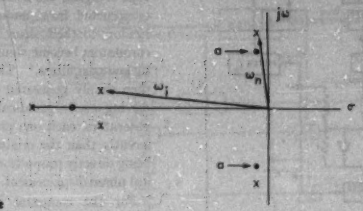
Executive-Control for Lightly Damped Varying Process

A common problem which is encountered in the design of practical control systems is one which involves process-transfer functions with one complex pole-pair close to the real frequency axis. Although the transfer function may contain additional poles and zeros as, for example, in Fig. 5(A), the behavior of the process when included in a closed loop is dominated by the poles adjacent to the $j\omega$ axis. In particular, with a relatively small loop gain, the system tends to become unstable.

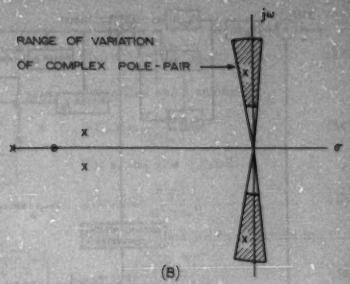
If the process is fixed, the usual approach to improving the closed-loop operation is to add a compensating network with zeros in close proximity to the poles, as for example, at position a in Fig. 5(A). If, however, the process poles vary in an unpredictable manner over a large region as illustrated in Fig. 5(B), it is extremely difficult, and sometimes impossible, to stabilize the system with fixed compensation without introducing undesirable effects into the over-all operation of the closed-loop system.

By applying the principles of executive-control, it is not only possible to stabilize the system, but it is also possible to maintain an approximately nonvarying, desirable, closed-loop transfer function in spite of large variations in the position of the offending pole-pair. In addition, for this problem, the executive-controller and the actuator are realizable as passive networks. This passive realization results in a relatively simple executive-controlled system.

In the actuator-controller of Fig. 3, the actuator transfer function is constructed to be $1/G(s)$, the inverse of the process-transfer function; and the executive-controller, which is responsible for achieving the specified operation, is a model of the desired forward-path transfer characteristic. For the problem under discussion, this approach is modified so that the actuator cancels only the effects due to the dominating pole-pair. Thus, if the process transfer function is:



(A)



(B)

Fig. 5. S-plane diagram of pole-pair variation

$$G(s)=\frac{N(s)}{D(s)}=\frac{N(s)}{D_1(s)[s^2+2\zeta\omega_n s+\omega_n^2]} \tag{5}$$

the actuator transfer function is constructed to be:

$$G_a(s)=\frac{[s^2+2\zeta\omega_n+\omega_n^2]}{D_a(s)} \tag{6}$$

where the poles of $G_a(s)$ are added to make $G_a(s)$ realizable with a linear passive network.

Since the dominant pole-pair changes position in the s-plane, it is necessary for $G_a(s)$ to be variable if the offending pole-pair is to be approximately cancelled at all times. Assuming that a plant observer exists which is capable of supplying the actuator with the approximate location of the important pole-pair, and assuming that $G_a(s)$ is able to vary accordingly, the transfer function of the executive-controller can be immediately determined as:

$$G_c(s)=G_d(s)\,\frac{D_1(s)D_a(s)}{N(s)} \tag{7}$$

A block diagram of the complete system is shown in Fig. 6(A).

Since it is not necessary to cancel exactly the dominant pole-pair, $G_a(s)$ need not be continuously variable. Instead, several fixed passive networks can be employed to cover effectively the region over which the pole-pair moves. Such an arrangement is shown in Fig. 6(B). The $G_a(s)$ networks are selected by a rotary switch. The function of the plant observer in such a system is to measure the approximate pole location and to choose the appropriate switch position.

If the region of pole variation closely surrounds the $j\omega$ axis, then ω_n, the resonant frequency of the dominant pole-pair, is sufficient for locating the pole position for the purposes of switching in the successive compensation networks. A plant observer capable of supplying the value of ω_n can be constructed by taking advantage of the properties of the frequency response characterization of a lightly damped process.

If, for the moment, it is assumed that in Fig. 5, ω_n is less than all other ω_i where the ω_i are the radial distances to all other poles and zeros of the process transfer function, then the frequency response of the process is of the form drawn in Fig. 7. If, in addition, ω_1 and ω_2 are chosen in such a manner that $\omega_1<\omega_n<\omega_2$ over the entire range of variation of ω_n, and where $\omega_2<\omega_i$, then it is shown in Appendix I that:

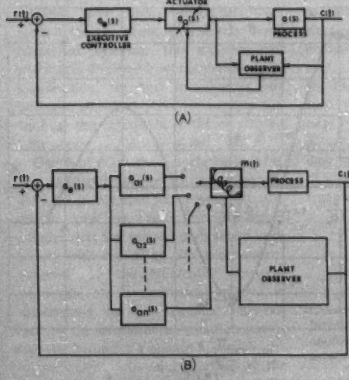
$$\omega_n=\omega_2\sqrt{\frac{|C(\omega_2)|}{|C(\omega_2)|+|KM(\omega_2)|}} \tag{8}$$

and

$$K=\left|\frac{C(\omega_1)}{M(\omega_1)}\right|=|G(\omega_1)| \tag{9}$$

where $C(\omega_1)$ and $C(\omega_2)$ are obtained from two narrow-band filters at the process output, and $M(\omega_1)$ qnd $M(\omega_2)$ are similarly obtained with two filters at the process input. ($M(\omega)$ and $C(\omega)$ are the Fourier transforms of $m(t)$ and $c(t)$, the process input and output.)

A straightforward analog computer



(A)

(B)

Fig. 6. Notch filter approach

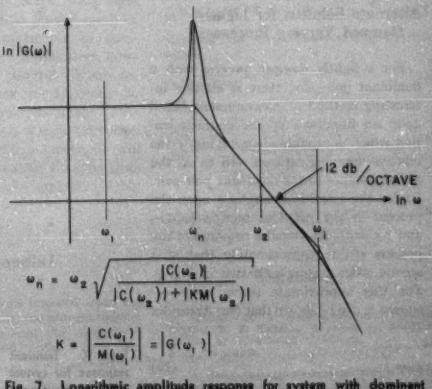


Fig. 7. Logarithmic amplitude response for system with dominant pole-pair

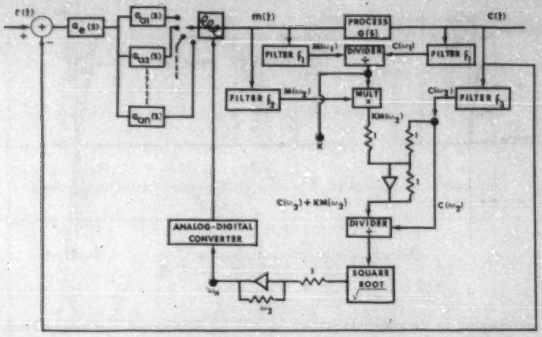


Fig. 8. Executive-controlled system diagram for frequency-response approach

can be used to solve equations 8 and 9. A complete executive-controlled process-adaptive system employing such a computer is illustrated in Fig. 8. It is interesting to note that in addition to the approximate pole position, the computer also supplies the value of $K = G(\omega_1)$, the low-frequency gain of the process. Thus, with slight modification, the system of Fig. 8 can be made adaptive to changes in gain as well as to changes in pole position.

If there is a pole at $s = 0$, the derivative of $c(t)$ is utilized by the plant observer instead of $c(t)$. If, after accounting for the poles at the origin there are one or more poles located so that it is impossible to choose ω_2 less than the smallest ω_i, equation 8 can be modified so that it will still give ω_n, provided the ω_i remain relatively fixed.

Alternate Solution for Lightly Damped Varying Processes

For a lightly damped process with a dominant pole-pair, there is another interesting method of measurement based upon an impulse response characterization which is capable of supplying, ζ the relative damping ratio as well as ω_n, the radial distance to the dominant pole-pair. Under the conditions imposed upon the process in the previous section ($\omega_n < \omega_i$ and ζ small), the impulse response of the process $g(t)$ is approximately that of a second-order system as drawn in Fig. 9. For this approximate second-order response it can be shown that (see Appendix II):

$$\omega_n = \frac{\pi}{T_0} \tag{10}$$

and that

$$\zeta = \frac{1}{\pi} \ln \left| \frac{g_M}{g_m} \right| \tag{11}$$

where T_0 is the period to the first zero of $g(t)$, and where g_M and g_m are the first maximum and first minimum respectively of $g(t)$ as indicated in Fig. 9.

A computer capable of determining ω_n and ζ from equations 10 and 11 is illustrated in Fig. 10. In contradistinction to the two previous methods, this approach utilizes test signals to aid in the measurement. If white noise is added to the process input, and if the process output is cross-correlated with the noise delayed by a constant τ, the output from the cross-correlator is $g(\tau)$, one point on the process impulse response.[6] There is a tremendous instrumentation advantage if, instead of ordinary noise, 2-level, or so-called binary, noise is employed. For this type of signal, the noise generators can be constructed from motor-driven cams or random-toothed discs,[7] and the cross-correlators become simple relays followed by low-pass filters.[8] The delayed noise is then simply generated by synchronously driving a set of identical mechanical noise generators, each one phased slightly differently than the original, the time delay being directly proportional to the mechanical phase displacement.

For the computer of Fig. 10, four mechanical noise generators (*NG*1 to *NG*4) are driven by motor no. 1 and separated from each other by mechanical phase shifters as indicated schematically in the diagram. The noise output from *NG*1 is added to $m(t)$, the process input. *NG*2 is connected to *NG*1 through a variable phase shifter (a simple differential gearing arrangement). The output of *NG*2 is is cross-correlated with the process output $c(t)$ thereby obtaining one point on the $g(\tau)$ curve. If the phase shift is initially zero, and if the relative phase between *NG*1 and *NG*2 is continuously changed by motor no. 2, the output from the first cross-correlator will scan the $g(\tau)$ curve. A null detector at this output can be employed to stop the motor when $g(\tau)$ goes through its first zero at $\tau = T_0$. Since τ is proportional to the angle of the variable phase shifter and, since ω_n is inversely proportional to T_0, the value of ω_n can be obtained directly from this first phase shifter.

If *NG*3 and *NG*4 are connected to *NG*2 through fixed gear ratios of 1/2 to 1 and 3/2 to 1 respectively, when *NG*2 is delayed by T_0, *NG*3 will be delayed by $T_0/2$ and

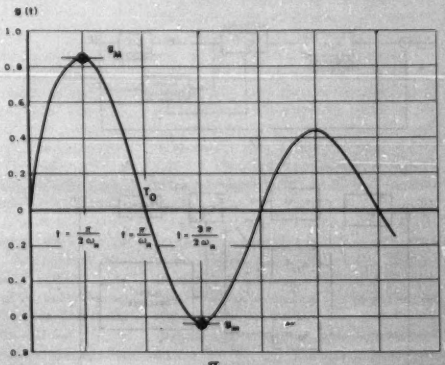


Fig. 9. Transient response for system with dominant pole pair

$$\omega_n = \frac{\pi}{T_0}$$

$$\zeta \cong \frac{1}{\pi} \ln \left| \frac{g_M}{g_m} \right|$$

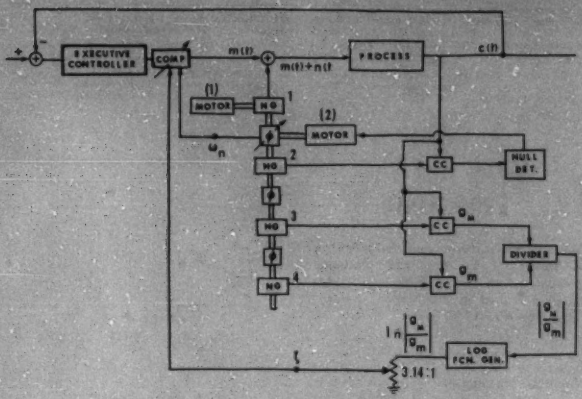


Fig. 10. Executive-control for impulse-response approach

*NG*4 by $3T_0/2$. But $T_0/2$ and $3T_0/2$ are approximately the values of τ at which $g(\tau)$ hits its first maximum and its first minimum. Thus, the outputs from the second and third cross-correlator are g_M and g_m. A divider followed by a logarithmic function generator and a potentiometer divider is sufficient to obtain ζ; see equation 11. The values of ω_n and ζ can then be used to adjust the compensator.

The compensator, once again, may be continuously or discretely variable. The total installation for this method is well within the limits of practical realizability. Since the emphasis in the last two sections has been on cancelling the effects of the offending process pole-pair, the resultant systems have been process-adaptive. Since in these systems, as in all executive-controlled systems, the measurement problem is divorced from the performance specification, the final installations can be made signal-adaptive by varying the transfer characteristics of the executive-controller.

One final point that should be made is that, although the performance specification can be changed instantaneously, the measurement procedures discussed herein require that the process parameters vary slowly with respect to the measurement interval. In the differential equation approach, the measurement interval is determined by the sampling intervals and by the simultaneous equation solver; for the frequency-response approach, the measurement interval is limited by the delay in the narrow-band filters; and for the transient-response method, the delay in the cross-correlators is the limiting factor.

Conclusions

Executive-control appears to be a useful approach for the design of adaptive systems. By dividing the responsibilities of the computer into three distinct functions: process measurement, performance maintenance, and signal generation, the problems associated with process adaptability and signal adaptability can be treated separately, with the final designs fused together to form a completely adaptive system.

The major challenge, at present, is in the process-measurement stage. Attempts to solve this problem in general result in large computer requirements of questionable practicality. Fortunately, however, for specific control situations, the total installations can apprently be kept well within practical limits. Perhaps from the solution of such specialized problems, some general methods will develop which will enable the principles of adaptive control to be applied to highly complex systems.

Appendix I

Under the conditions imposed upon the process in Figs. 5 and 7, the process-transfer function for $\omega<\omega_t$ can be considered approximately second order. Thus, for $\omega<\omega_t$:

$$G(s)=\frac{K\omega_n^2}{(s^2+2\zeta\omega_n s+\omega_n^2)} \tag{12}$$

or in terms of ω,

$$G(\omega)=\frac{K\omega_n^2}{(\omega_n^2-\omega^2)+2j\zeta\omega_n\omega} \tag{13}$$

For $\omega_1 \ll \omega_n$ and ζ small,

$$|G(\omega_1)|=\left|\frac{C(\omega_1)}{M(\omega_1)}\right|\cong K \tag{14}$$

For $\omega_2 \gg \omega_n$ and ζ small,

$$|G(\omega_2)|\cong\frac{K\omega_n^2}{(\omega_2^2-\omega_n^2)} \tag{15}$$

Solving equation 15 for ω_n results in:

$$\omega_n=\omega_2\sqrt{\frac{|G(\omega_2)|}{|G(\omega_2)|+K}} \tag{16}$$

but

$$|G(\omega_2)|=\left|\frac{C(\omega_2)}{M(\omega_2)}\right| \tag{17}$$

Thus,

$$\omega_n=\omega_2\sqrt{\frac{|C(\omega_2)|}{|C(\omega_2)|+K|M(\omega_2)|}} \tag{18}$$

Appendix II

Once again, if $G(s)$ is assumed approximately second order:

$$g(t)=\frac{K\omega_n}{\sqrt{1-\zeta^2}}\,e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t\right) \tag{19}$$

where $g(t)$ is the process-impulse response. If the process is lightly damped, ζ is small and

$$g(t)\cong K\omega_n e^{-\zeta\omega_n t}\sin\omega_n t \tag{20}$$

The first time $g(t)$ crosses zero (excluding $t=0$) is at $t=\pi/\omega_n=T_0$. Therefore,

$$\omega_n=\frac{\pi}{T_0} \tag{21}$$

But at $t=\pi/2\omega_n=T_0/2$, $g(t)=g_M$. Hence,

$$g_M=K\omega_n e^{-\pi/2} \tag{22}$$

However, at $t=3\pi/2\omega_n=3T_0/2$, $g(t)=g_m$ where

$$g_m=-K\omega_n e^{-3\pi/2} \tag{23}$$

Therefore,

$$\left|\frac{g_M}{g_m}\right|=e^{\pi\zeta} \tag{24}$$

or

$$\zeta\cong\frac{1}{\pi}\ln\left|\frac{g_M}{g_m}\right| \tag{25}$$

References

1. Design of a Self-Optimizing Control System, R. E. Kalman. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., Feb. 1958.

2. A Self-Adjusting System for Optimum Dynamic Performance, G. W. Anderson, J. A. Aseltine. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1958.

3. Adaptive Servomechanisms, R. F. Dronick, R. A. Shahbender. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, Nov. 1957, pp. 286–92.

4. Executive-Controlled Adaptive Systems, R. Staffin. *Dissertation*, Doctorate, Electrical Engineering, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., June 1959.

5. Executive-Controlled Adaptive Systems, R. Staffin, J. G. Truxal. *Research Report R-632-58, PIB-616*, Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., Sept. 9, 1958.

6. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

7. Dynamic Response Measurement in the Presence of Noise and Limiting, E. S. Joline. *M.S. Thesis*, Polytechnic Institute of Brooklyn, June 1955.

8. A Survey of Adaptive Control Systems, J. A. Aseltine, A. R. Mancini, C. W. Sarture. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, Dec. 1958.

ERRATUM

"Some Effects of Hypersonic Ionization on the Design of Electrical and Electronic Components" by W. B. Sisco and J. M. Fiskin; published in Applications and Industry, November 1959, pages 352–56.

The authors wish to make the following corrections in their paper. On page 356, column 2, in the first paragraph the allusion to "equations 2, 3, and 4" should read "equations 1, 2, and 3." Also, items 1 and 12 of the "References" should be interchanged.